2001

Structural genomics of the barley Mla powdery mildew resistance complex

Fusheng Wei
Iowa State University

Structural genomics of the barley *Mla* powdery mildew resistance complex

By

Fusheng Wei

A dissertation submitted to the graduate faculty
In partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Major: Genetics
Major Professor: Roger P. Wise

Iowa State University
Ames, Iowa
2001

Graduate College

Iowa State University

This is to certify that the Doctoral dissertation of

Fusheng Wei

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

Signature was redacted for privacy.

For the Graduate College

TABLE OF CONTENTS

CHAPTER 1. GENERAL INTRODUCTION

Overview

Barley (*Hordeum vulgare* L.) is one of the earliest domesticated species and is the fourth major grain product in the world. The short growing season, many wild resources, broad geographical distribution, and great adaptability make barley a good model plant for research in large-genome small grain plants.

Barley powdery mildew, caused by the biotrophic fungus *Blumeria* (*Erysiphe*) *graminis* f. sp. *hordei*, has long been one of the most important diseases in the world. Resistance to powdery mildew in barley is conferred by genes designated *Ml*, which are widely distributed in different chromosomes (Jorgensen, 1994). For the *Mla* locus, over 30 alleles or tightly linked genes have been identified (Giese, 1981; Giese et al., 1981; Wise and Ellingboe, 1985; Jahoor and Fischbeck, 1993). Another intriguing feature of this system is that there are both race- and non-race specific resistances in barley. The race-specific *Ml* genes are dominant while non-race specific *mlo* is recessive. Extensive studies on genetics, cytology and physiology of barley powdery mildew have been conducted (Bryngelsson and Collinge, 1992; Scott, 1992). Thus, the barley powdery mildew is an excellent system to study plant-pathogen interaction.

To structurally characterize the barley *Mla* cluster, high-resolution genetic and physical maps are essential. A high-resolution inbred mapping population was previously established (Mahadevappa et al., 1994). The AFLP (amplified fragment length polymorphism) technique (Vos et al., 1995) in conjunction with bulk segregation analysis (Michelmore, et al., 1991) were introduced to saturate the *Mla* region. A physical map spanning the *Mla* locus was built by using the tightly-linked genetic markers to screen large-insert YAC (yeast artificial chromosome) and BAC (bacterial artificial chromosome) libraries. The end

sequences of these large-insert clones were utilized to do further genetic and physical walking on the *Mla* region. A low-pass sequencing strategy was used to estimate the gene content of the *Mla*-spanning clones.

To fully understand the gene content and genome structure of the *Mla* cluster, we adopted the newly-developed high throughput sequencing approach. Computational analysis was employed to annotate the assembled fragment and biological function was assessed.

Dissertation Organization

This dissertation contains three research manuscripts and a general conclusion. The first manuscript was published in Genetics (153: 1929-1948, 1999) and describes genetic and physical mapping of the barley *Mla* powdery mildew resistance locus. The *Mla* candidate genes are also identified in this manuscript. The second manuscript has been submitted to Nature Genetics and describes the gene content, genome organization and molecular evolution of the barley *Mla* resistance complex. The third manuscript is being prepared for Genome Research and describes the molecular cloning of a heterochromatic knob-like sequence and genes within this gene-rich region are transcriptionally active. The final part of this dissertation summarizes our results and recommends the future direction of this study.

Literature Review

The interaction between plant and its surrounding organisms could be described as coexisting (none is hurt) and competing (one is hurt). The relationship between a pathogen and its host belong to the latter. Plants have developed many mechanisms to enable them to survive in their environments. One mechanism is the gene-for-gene interaction (Flor 1956), in which the product of a host resistance (*R*) gene interacts with the product of a pathogen avirulence (*Avr*) gene and thus activates the signal transduction pathway of active defense.

To date, forty *R* genes have been cloned (Table 1). These cloned *R* genes confer resistance to a wide range of pathogens including viruses, bacteria, fungi, nematodes and insects. Most of these genes exhibit characteristics that suggest their involvement in a signal transduction pathway. According to their functional domain, most of the cloned *R* genes fall into the NBS-LRR (nucleotide binding site, leucine-rich repeat) class. The difference in the N terminal motifs of the *R* gene products in the NBS-LRR class further divides them into two subclasses, TIR (Toll/interleukin receptor)-NBS-LRR and CC (coiled-coil) or LZ (leucine zipper)-NBS-LRR. The classification of these *R* genes has not recently been updated. In the non-TIR subclass of NBS-LRR disease resistance genes, there is still some confusion whether an *R* gene belongs to the leucine zipper, CC, or neither subclass. In this review, the classification of those *R* genes is clarified and their potential role in the signal transduction pathway is postulated.

Classification of cloned disease resistance genes

R genes can be classified into eight classes according to the conserved domains in their putative protein products. Of the 40 cloned *R* genes (listed in Table 1), thirty-four have LRR and two have LRD (Leucine-Rich Domain), together designated as LRR(D). The up- or down-stream conserved domains of these thirty-six LRR(D) *R* genes further divide them into four classes, designated NBS-LRR(D), LRR(D), LRR-TM (transmembrane), and LRR-TM-kinase classes. The remaining four *R* genes could be grouped into four classes with a single member in each class.

The LRR domains function to mediate protein-protein, protein-ligand, and protein-carbohydrate interactions (Kobe and Deisenhofer, 1995). Within a single LRR motif in the LRR domain, the xxLxLxx motif folds as a ß-sheet in which the leucine residues (L) form a hydrophobic core, whereas the side-chains of the flanking amino acid residues (x) are solvent exposed. In multiple LRR motifs in the LRR domain, the ß-sheets are aligned in parallel and

Table 1. Classification of cloned disease resistance gene

Class	Gene	GeneBank Protein ID #	Host	Pathogen	Reference
HC toxin reductase	*Hm1*	T03970	maize	*Cochliobolus carbonum*	Johal and Briggs, 1992
CC	*RPW8*	AF273059	*Arabidopsis*	*Erysiphe cichoracearum*	Xiao et al., 2001
NBS-LRR(D)					
TIR-(CC)NBS_1-LRR	*L*	AAA91022	flax	*Melampsora lini*	Lawrence et al., 1995
	M	AAB47618	flax	*Melampsora lini*	Anderson et al., 1997
	P		flax	*Melampsora lini*	Dodds et al.,s 2001
	N	A54810	tobacco	*Tobacco mosaic virus*	Whitham et al., 1994
	RPP5	AAF08790	*Arabidopsis*	*Peronospora parasitica*	Noel et al., 1999
	RPP1	AAC72977	*Arabidopsis*	*Peronospora parasitica*	Botella et al., 1998
		AAC72978			
		AAC72799			
	RPS4	T51140	*Arabidopsis*	*Pseudomonas syringae*	Gassmann et al., 1999
CC-NBS_2-LRR(D)	*RPS2*	A54809	*Arabidopsis*	*Pseudomonas syringae*	Bent et al., 1994
	RPS5	AAC26126	*Arabidopsis*	*Pseudomonas syringae*	Warren et al., 1998
	RPP8	T48898	*Arabidopsis*	*Peronospora parasitica*	McDowell et al., 1998
	RPP13	AAF42832	*Arabidopsis*	*Peronospora parasitica*	Bittner-Eddy et al., 2000
	HRT	AAF36987	*Arabidopsis*	*Turnip crinkle virus*	Cooley et al., 2000
	Mi	AAC97933	tomato	*Meloidogyne incognita*	Milligan et al., 1998
				Macrosiphum euphorbiae	Vos et al., 1998
					Rossi et al., 1998
	Sw-5		tomato	*tospovirus*	Brommonschenkel et al., 2000

Table 1. (continued)

Class	Gene	GeneBank Protein ID #	Host	Pathogen	Reference
CC-NBS$_2$-LRR(D)	*I2*	AAD27815	tomato	*Fusarium oxysporum*	Ori et al., 1997 Simons et al., 1998
	Prf	T07589	tomato	*Pseudomonas syringae*	Salmeron et al., 1996
	Dm3	AAD03156	lettuce	*Bremia lactucae*	Meyers et al., 1998
	Mla1	AAG37356	barley	*Blumeria graminis*	Zhou et al., 2000
	Mla6	CAC29242	barley	*Blumeria graminis*	Halterman et al. 2001
	Pi-ta	AAK00132	rice	*Magnaporthe grisea*	Bryan et al., 2000
	Xa1	T00020	rice	*Xanthomonas oryzae*	Yoshimura et al., 1998
	Pib	BAA76282	rice	*Magnaporthe grisea*	Wang et al., 1999
	Rp1D	AAD47197	maize	*Puccinia sorghi*	Collins et al., 1999
NC-NBS$_2$-LRR	*RPM1*	A57072	*Arabidopsis*	*Pseudomonas syringae*	Grant et al., 1995
	Rx	CAB50786	potato	*potato virus X*	Bendahmane et al., 1999
	Gpa2	AAF04603	potato	*Globodera pallida*	Van Der Vossen et al., 2000
	Bs2	AAF09256	pepper	*Xanthomonas campestris*	Tai et al., 1999
LRR(D)	*Hs1^{Pro1}*	AAB48305	sugarbeet	*Heterodera schachtii*	Cai et al., 1997
	Xa21D	AAB82753	rice	*Xanthomonas oryzae*	Wang et al., 1998
LRR-TM	*Cf2*	T10504 T10515	tomato	*Cladosporium fulvum*	Dixon et al., 1996
	Cf5	AAC78591	tomato	*Cladosporium fulvum*	Dixon et al., 1998
	Cf4	CAA05268	tomato	*Cladosporium fulvum*	Parniske et al. 1997
	Cf9	CAA05274	tomato	*Cladosporium fulvum*	Jones et al. 1994

Table 1. (continued)

Class	Gene	GeneBank Protein ID #	Host	Pathogen	Reference
LRR-TM	*Cf-4A*	T07015	tomato	*Cladosporium fulvum*	Takken et al., 1998
	Hcr9-4E	CAA05269	tomato	*Cladosporium fulvum*	Takken et al., 1999
LRR-TM-kinase	*Xa21*	AAB82753	rice	*Xanthomonas oryzae*	Song et al., 1995
TM	*mlo*	P93766	barley	*Blumeria graminis*	Büschges et al., 1997
Kinase	*Pto*	A49332	tomato	*Pseudomonas syringae*	Martin et al., 1993

form a surface decorated with solvent-exposed residues that can interact specifically with a ligand (Kobe and Deisenhofer, 1995). The substantial sequence variation among members of an *R* gene family indicates that the LRR domain of *R* genes plays a direct role in determining the specificity in gene-for-gene interactions (Parniske et al. 1997; McDowell et al. 1998; Meyers et al. 1998; Ellis et al. 1999). The LRR variation in flax *L* alleles, which confer resistance to *Melampsora lini*, and in tomato *Cf* genes, which confer resistance to *Cladosporium fulvum*, strongly suggests alternate specificities for pathogen recognition corresponding to the LRR variation (Parniske et al. 1999; Ellis et al. 1999). Recently Jia et al. (2000) have shown the direct interaction of AVR-Pita$_{176}$ and Pi-ta LRD region inside the plant cell to initiate a Pi-ta-mediated defense response to the rice blast pathogen *Magnaporthe grisea*.

LRR domains also function in downstream signaling in addition to their primary function in AVR protein recognition. For example, a glutamate-to-lysine substitution in the C-terminal LRR alters the specificity of *RPS5* to *Pseudomonas syringae*. This substitution also partially suppresses other *R* genes from functioning, such as *RPM1* and *RPS2*, both resistant to *Pseudomonas syringae*, and *RPP2, RPP4, RPP6*, or *RPP9*, all resistant to downy mildew pathogen *Peronospora parasitica* (Warren et al. 1998).

The most prevalent class of cloned resistance genes contains both NBS and LRR(D) domains. The NBS domain contains various motifs, designated P-loop, kinase 2, and kinase 3a. These motifs can be found in many ATP- and GTP-binding proteins including the proteins involved in the animal Ras signal pathway (together called RAS group), ATPases, elongation factors, and G-proteins (Saraste et al. 1990). The NBS domain of plant disease resistance genes is similar to that of CED4/APAF1 from *C. elegans* and human, and thus forms the NB-ARC (for nucleotide binding APAF1-resistance protein CED4) domain (van der Biezen et al. 1998). The characteristic of molecular switches in NBS proteins are

important in many eukaryotic cellular events such as cell growth, differentiation, cytoskeletal organization, vesicle transport, apoptosis, and defense (van der Biezen et al. 1998; Aravind et al. 1999; Bourne et al. 1991). The P-loop motif, [GXXGXGKTT in *R* genes. GXXXXGK(T/S) in general] has been reported to interact with phosphates and Mg^{2+} ions (Saraste et al. 1990). The kinase 2 motif (LIVLDDV) is believed to function in a phospho-transfer reaction with the extreme conserved first aspartate amino acid coordinating the divalent metal ion (Mg^{2+}). The kinase 3a motif is involved in binding purine or ribose and contains a tyrosine or arginine. However, to date there is no biochemical evidence to substantiate the nucleotide-binding role of the NBS of resistance gene products. The NBS of Ced-4 and its human ortholog Apaf-1 have been reported to execute apoptotic cell death (Li et al. 1997). The apoptotic-like hypersensitive responses of disease resistance in plants suggest the possibility of a similar pathway for the two systems.

Sequence alignment and phylogenetic analysis of the NBS domain in the NBS-LRR class of disease resistance genes showed that the NBS domain falls into two distinct groups (Pan et al. 2000). Group I is associated with N terminal TIR domain and Group II associated with the N terminal CC domain. To distinguish these two groups of NBS domains, Group I NBS domain is designated as NBS_1 whereas Group II as NBS_2. The classification of the Group I TIR-NBS_1-LRR subclass is very consistent, but, Group II is not. For example, in *RPM1*, *Rx/Gpa2* and *Bs2* families, the possibility of having an N-terminal CC domain is low, between 28 to 38%, although their NBS domains fit well to the CC-NBS_2-LRR subclass. Meanwhile, a database search for conserved domain (CD) did not reveal the LZ motif in all the NBS-LRR class of disease resistance genes even though the LZ motif of R gene products was originally postulated from the *Arabidopsis* RPM1 protein. Therefore, it is reasonable to divide the NBS_2-LRR subclass into two subgroups, CC-NBS_2-LRR and NC-NBS_2-LRR (No-

CC- NBS_2-LRR). The CC-NBS_2-LRR subclass is the largest among the NBS-LRR(D) class of *R* gene (Table 1).

The sharing of the TIR domain in the NBS_1-LRR subclass of *R* gene with Drosophila Toll or human interleukin receptor may indicate common origins for innate resistance in multicellular organisms (Baker et al. 1997; Hammond-Kosack and Jones 1997). The Toll receptor in Drosophila is essential for the dorsoventral pattern development in embryos and the immune response in the adult fly. Human orthologues of the Toll protein have been shown to signal adaptive immunity via NF-kappaB and mediate lipopolysaccharide-induced cellular signaling (Medzhitov et al. 1997; Yang et al. 1998). This suggests the involvement of Toll homologues in pathogen recognition in different multicellular organisms.

As one of the most common protein structure motifs, CC domains mediate protein–protein interactions (Lupas, 1996). These protein–protein interactions have been found to be required for binding to another protein and thus for their function (Cheng et al. 2001). The leucine zipper domain in some transcription factors is the best example of CC domain (Landschulz et al., 1988; McKnight, 1991; Ellenberger et al., 1992).

Both the TIR domain and the NBS_1 domain have only been found in dicotyledenous plants (Pan et al. 1999). More surprisingly, when we analyzed the NBS_1-LRR subclass of *R* genes, we found a CC domain embedded in all members of this subclass (Table 1) although the positions of CC domains may vary in the NBS_1 domain. To show this characteristic of the embedded CC domain, the NBS_1-LRR subclass is designated as TIR-(CC)NBS_1-LRR. The presence of CC domain in the NBS_1-LRR subclass of *R* genes strongly suggests that protein oligomerization is essential for *R* gene function. Recent reports have shown that these domains are essential for disease resistance. For example, mutations in the TIR, NBS and LRR domains for the tobacco *N* gene and in the LZ, NBS, and LRR for the *Arabidopsis*

RPS2 gene abolished the two *R* gene functions respectively (Dinesh-Kumar et al. 2000; Tao et al. 2000).

Hs1^{pro-1} and *Xa21D* belong to the LRR class of resistance genes (Table 1; Cai et al. 1997; Wang et al. 1998). The *Hs1*$^{pro-1}$ gene confers resistance to the beet cyst nematode (*Heterodera schachtii* Schmidt), a major pest in sugar beet (*Beta vulgaris* L.). The rice *Xa21D* gene confers resistance to *Xanthomonas oryzae* pv *oryzae* in a race-specific manner. The LRRs of predicted Hs1^{pro-1} and XA21D protein are presumed extracellular.

Cf genes belong to the single family of LRR-TM class of resistance genes (Table 1; Dixon et al. 1996; 1998; Jones et al. 1997; Parniske et al. 1997; Takken et al. 1998; 1999). *Cf* genes encode extracellular LRR proteins with single membrane spanning regions and short cytoplasmic carboxyl termini. The *Cf* genes from tomato (*Lycopersicon esculentum*) confer resistance to the fungal pathogen *Cladosporium fulvum*, a biotrophic extracellular fungus without haustoria that attacks the host leaf.

The lone member of LRR-TM-kinase class of resistance genes is *Xa21*, which confers resistance to multiple *Xanthomonas oryzae* pv. *oryzae* isolates in rice (Table 1; Song et al. 1995). The predicted XA21 protein carries an extracellular LRR, a transmembrane domain and a cytoplasmic serine-threonine kinase-like domain. These domains suggest that XA21 plays a role in cell surface recognition of a pathogen ligand and subsequent activation of an intracellular defense response.

The *Pto* kinase is the only member of the serine-threonine kinase class of resistance genes (Table 1; Martin et al. 1995). The *Pto* gene confers resistance to *Pseudomonas syringae* pv. *tomato* (*Pst*) in tomato. *Pto* requires a CC-NBS_2-LRR gene, *Prf* for its function (Salmeron et al. 1996). Overexpression of *Pto* and *Prf* confers broad-spectrum resistance in tomato (Oldroyd and Staskawicz, 1998; Rathjen et al. 1999; Tang et al. 1999). The *Pto*

product can directly bind to the products of *Ptil, 4, 5,* and *6* and then initiates HR and defense response (Zhou et al. 1997; Gu et al. 2000).

The barley *mlo* gene is the only member of the TM class of *R* genes (Table 1; Büschges et al. 1997). The recessive *mlo* is resistant to all races of the powdery mildew fungus *Blumeria graminis f.* sp. *hordei.* MLO is membrane-anchored by 7 transmembrane (TM) helices such that the N terminus is located extracellularly and the C terminus intracellularly (Devoto et al. 1999). Plants that contain the *mlo* gene show spontaneous lesions without pathogen infection. The dominant *Mlo* gene has been proposed to perform a dual negative control function in the cell death and the onset of disease resistance.

The *Arabidopsis RPW8* gene, conferring resistance to all the pathogens of *Arabidopsis* powdery mildew, is the single member of CC class of resistance gene (Table 1; Xiao et al. 2001). *RPW8* contains two naturally polymorphic, dominant and functional *R* genes, *RPW8.1* and *RPW8.2*. These two *RPW8* genes induce localized, salicylic acid-dependent defenses similar to those of race-specific *R* genes. Thus, it suggests that broad-spectrum resistance mediated by *RPW8* uses the same mechanisms as race-specific resistance.

The HC toxin reductase class of disease resistance gene was the first disease resistance gene to be cloned (Table 1; Johal and Briggs, 1992). The maize *Hm1* gene is the sole member of this class. *Hm1* encodes an NADPH-dependent HC toxin reductase and confers resistance to the leaf spot fungus *Cochliobolus carbonum* race 1. The HM1 reductase enzyme detoxifies the *C. carbonum* HC-toxin. This toxin inhibits histone deacetylase activity.

An integrated signaling pathway for disease resistance genes

The cloning of disease resistance genes gives us an opportunity to further understand the functional signaling pathway of these genes. The conserved domains of *R* genes indicate that different *R* genes function in similar signal transduction patterns. Physiological and

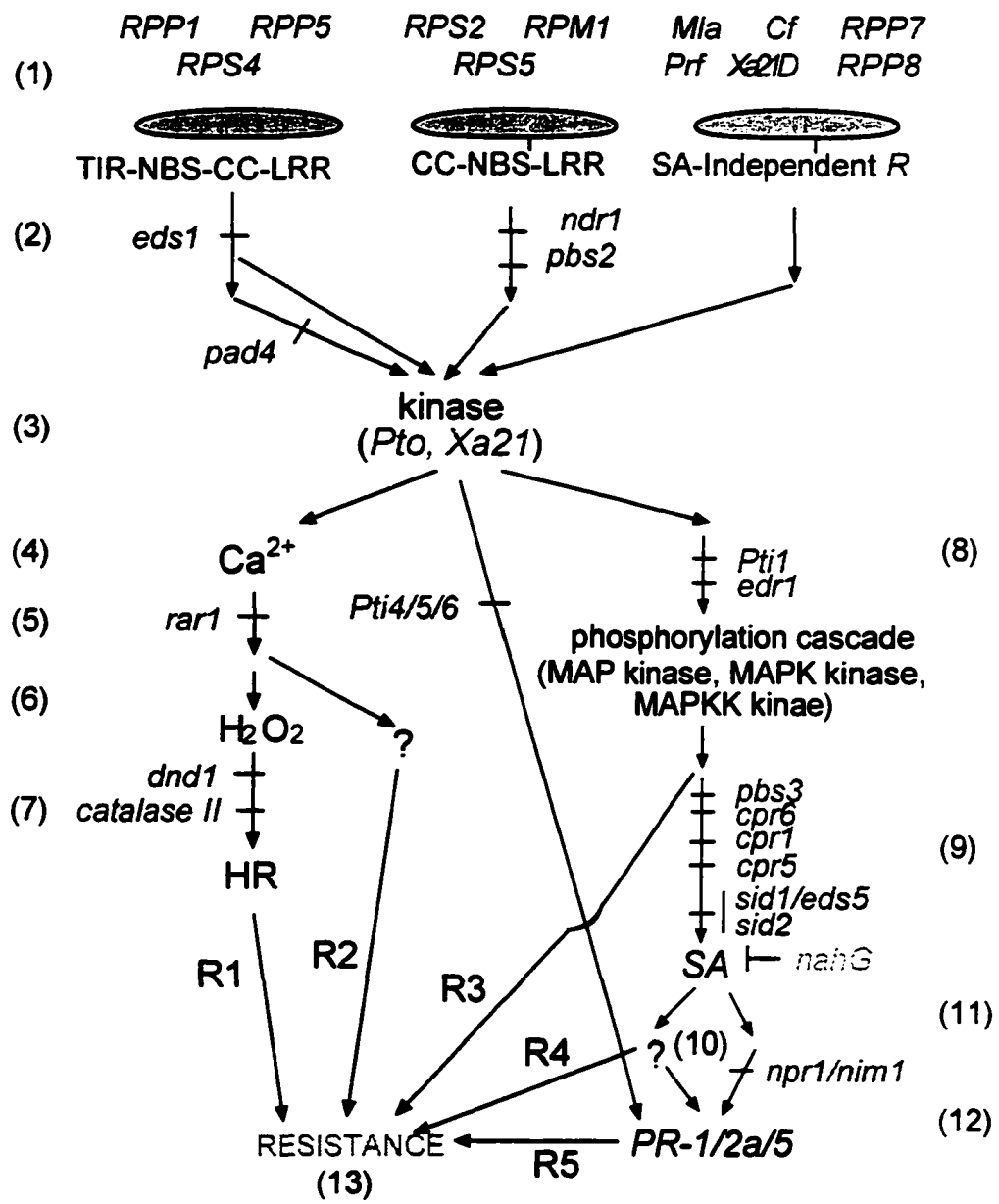
mutational analyses of *R-Avr* interactions have shown some functional convergence and divergence points for signaling. These convergence/divergence points could be metabolic products, such as salicylic acid (SA) and H_2O_2, or gene products, such as EDS1, NDR, and PR proteins. A phenotype, such as hypersensitive response (HR) or a biochemical cascade, such as MAP phosphorylation cascade could also contributes to such a convergence/divergence point. By comparing the components required for plant disease resistance with components involved in T cell receptor signaling in mammals, an integrated map of plant *R* gene signaling was constructed (Figure 1).

Early signal perception of R-Avr interaction

In *Arabidopsis*, the *EDS1* gene is required for the function of the TIR-NBS_1-LRR subclass of resistance genes, but not for CC-NBS_2-LRR type *R* genes. In contrast, the *NDR1* (Century et al. 1997) and *PBS2* genes are essential for the function of CC-NBS_2-LRR *R* genes, but not for TIR-NBS_1-LRR *R* genes. Mutations in *PAD4*, which is SA-dependent, cause some, but not all, TIR-NBS_1-LRR *R* genes to lose function, thus indicating that some TIR-NBS_1-LRR *R* genes are EDS1-dependent, but SA-independent. *EDS1*, *PAD4*, *NDR1* and *PBS2* all operate upstream of SA. Meanwhile there are some CC-NBS_2-LRR *R* genes such as *RPP7* and *RPP8*, that are SA- and *NDR*-independent (Aarts et al. 1998; Glazebrook, 1999; Feys and Parker, 2000). Barley *Mla* genes (Huckelhoven et al. 1999) as well as tomato *Prf*, *Pto* (Thara et al. 1999) and *Cf* (Brading et al. 2000) genes do not require SA for function, despite the fact that they reside in different *R* gene classes (Fig. 1.1 and 1.2). These SA-independent *R* genes are in at least another early signaling perception pathway.

The R-Avr interaction has been suggested to initiate a downstream signal transduction pathway. In mammalian systems, the perception of an antigen on the major histocompatibility complex (MHC) by T cell receptors (TCR) invokes kinase-mediated

Fig. 1. An integrated signaling pathway for plant disease resistance. At top of the figure are three groups of plant disease resistance genes, which trigger defense response after perceiving avirulence (*Avr*) gene products from pathogens. The different requirements of EDS1, NDR1 proteins and salicylic acid (SA) divide plant *R* genes into these three groups. NDR1 is required for some, not all CC-NBS-LRR *R* gene to function. Arrows indicate the direction of signal transduction and point to divergence/convergence points. The short horizontal lines show that a mutant gene was identified in that step. The numbers in the quotes show respective signaling steps. R1 to R5 indicate the respective resistance pathway.



edr: enhanced disease resistance
ndr: non race-specific disease resistance
dnd: defense, no death
cpr: constitutive expresser of PR genes
npr: non-expresser of PR
eds: enhanced disease suscetibility
pbs: Pseudomonas avrPphB susceptible
Pti: Pto interactor
pad: phytoalexin deficient
sid: salicylic acid induction-deficient
nah: enzymes responsible for conversion of naphthalene to 2-hydroxymuconic acid
rar: required for *Mla* resistance

phosphorylations and signal activation. These signals include calcium mobilization and activation of the Ras pathway. In plants, both Pto and XA21 are active kinases, and thus possibly play a similar role as activation agents (Fig. 1.3).

Calcium mobilization

EDS1 is a member of the lipase 3 (triacylglycerol lipase) family. It is interesting to note that mammalian TCR signaling is stimulated by a lipase, phospholipase Cgamma1 (PLCgamma1; Lin and Weiss, 2001). Activated PLCgamma1 cleaves phosphatidylinositol 4,5 bisphosphate [PtdIns(4,5)P2] at the plasma membrane and produces diacylglycerol (DAG) and inositol 1,4,5-triphosphate [Ins(1,4,5)P3]. As two essential components for T cell activation, DAG activates a number of proteins, such as the various isoforms of protein kinase C (PKC) and Ras guanyl-nucleotide-releasing protein (RasGRP), whereas Ins(1,4,5)P3 binds to Ins(1,4,5)P3 receptors (Ins(1,4,5)P3-Rs) on the surface of the endoplasmic reticulum (ER). The binding of Ins(1,4,5)P3 receptor with Ins(1,4,5)P3 helps Ca^{2+} stored in the ER be released into the cytoplasm, triggering the opening of Ca^{2+} -release-activated Ca^{2+} (CRAC) channels at the plasma membrane, and allowing influx of extracellular Ca^{2+}. The increased Ca^{2+} levels disrupt the inhibitory effects of calmodulin and then activate the protein phosphatase calcineurin. Calcineurin activation leads to the dephosphorylation of nuclear factor of activated T cells (NFAT), allowing it to enter the nucleus, where it cooperates with other transcription factors to bind promoters (Lin and Weiss, 2001).

Direct evidence of calcium involvement in the plant resistance comes from *Arabidopsis avrRpm1/RPM1* interaction (Grant et al. 2000) and analysis of calmodulin isoform proteins in pathogen induction and in transgenic plants (Levine et al. 1996; Harding et al. 1997; Heo et al. 1999). These analyses showed that calcium functions upstream of H_2O_2 accumulation and HR, and could be SA-independent. The Ca^{2+} transportation is often coupled with the

in/efflux of other ion such as K^+, H^+ and other metal ions (Fig. 1.4; Blumwald, et al. 1998; Hirschi, et al. 2001). Our recent review of functional clustering of defense related genes indicate that the barley *Rar1* and *Tip* (tonoplast intrinsic protein) perhaps function in the same pathway (Wei et al. 2001; Chapter 3 of this dissertation). The ion transporter in the tonoplast functions to regulate cytoplasmic ion levels and modulates the cytosolic level of Ca^{2+} in the signal transduction process. Therefore, it is reasonable to put the Zn^{2+}-binding protein RAR1 downstream of Ca^{2+} regulation (Fig. 1.5). The *Rar1* gene is required for the second oxidative burst and for resistance in the SA-independent, MLA-AvrMla interaction with HR. However, mutations in *Arabidopsis Dnd1* gene, a Ca^{2+}-permeable and cAMP-activated cyclic nucleotide-gated cation channel, abolish HR and yet still retain the gene-for-gene resistance (Clough et al. 2000). This separation of HR from resistance was also observed for potato *Rx*, a resistance gene to potato virus X, and for the oat *Pca* (crown-rust resistance) locus (Bendahmane et al. 1999; Yu and Wise, unpublished data). Therefore *Rar1* could be a divergence point for HR and disease resistance (Fig. 1.5).

Catalase is a family of tetrameric, heme-binding proteins that convert H_2O_2 to H_2O and O_2. H_2O_2 is produced during the oxidative burst in the onset of R-Avr interaction. The direct interaction between Xa21 and catalase B suggests the involement of catalase B in the *Xa21* signal pathway (Fig 1.7; Chern et al. 1999).

Activation of plant Ras-like pathway

In mammals, the recruitment of exchange factors Sos and RasGRP to the membrane activates the Ras pathway. The binding of the C1 domain of RasGRP to DAG is required for RasGRP function. Besides RasGRP, other proteins also may be involved in the Ras activation. A number of serine/threonine kinases and dual-specificity kinases are activated by the GTP-bound Ras. The dual-specificity kinases further activate the mitogen-activated protein (MAP) kinases Erk1/2, JNK and p38. These MAP kinases directly phosphorylate

transcription factors involved in the formation of the heterodimeric transcription factor AP-1. NF-kappaB is another important transcription factor for the generation of IL-2. Activation of NF-kappaB was involved in the stimulation of the TCR. The phosphorylation by the serine/threonine kinase Akt and the MAP kinase kinase kinases (MAPKKKs) activate the heterotrimeric IkappaB kinase complex. The IkappaB kinase (IKK) complex regulates NF-kappaB activity by phosphorylating IkappaB, which leads to its ubiquitination. Disassociated from IkappaB, NF-kappaB could move into the nucleus and activate transcription (Lin and Weiss, 2001).

In a similar fashion to the mammalian system, the Pti1 serine/threonine kinase can directly interact with Pto and is phosphorylated by Pto kinase. The Pti1 kinase also possibly participates in the phosphorylation cascade (Zhou et al. 1995; Bogdanove and Martin 2000). EDR1 is a MAP kinase kinase kinase protein and functions at the top of a MAP kinase cascade that negatively regulates SA-inducible defense responses (Frye et al. 2001). MAP kinase 4 is another kinase protein involved in the negative regulation of systemic acquired resistance in *Arabidopsis* (Petersen et al. 2000). Both SA-dependent *R* genes, such as the tobacco *N* gene and SA-independent *R* genes, such as the tomato *Cf-9* (Brading et al. 2000) require MAP phosphorylation for their functions (Zhang and Klessig, 1998; Remeis et al. 2000). The finding of resistance gene-dependent activation of a calcium-dependent protein kinase (Remeis et al. 1999) indicates that calcium regulation and phosphorylation cascade could be coupled.

SA-dependent downstream pathway

Many *R* genes in *Arabidopsis* have been shown to be SA-dependent by *nahG* transgenic plants, which can convert SA into hydroxymuconic acid immediately after its synthesis (the left two groups in Fig. 1.1). The *cpr* mutants constitutively express PR proteins and confer resistance to *Pseudomona syringae* and *Peronospora parastica* via an SA-dependent

pathway (Bowling et al. 1994; 1997; Clarke et al. 1998; 2000). All the three CPR genes act upstream of SA. The *cpr1* gene is proposed to negatively regulate SA-dependent resistance. The *pbs3* plants, which have reduced levels of SA, partially suppress resistance, thus indicating involvement in the SA pathway (Warren et al. 1999). For the *sid* genes, which make plants susceptible to both *Pseudomonas syringae* and *Peronospora parasitica*, the *nahG* analysis indicates their involvement in the SA synthesis (Fig. 1.9; Nawrath et al. 1999; Rogers and Ausubel, 1997). The *NahG* transgenic analysis and *edr1/npr1* double mutants indicated that the *edr* gene is SA and NPR1/NIM1 dependent. The PR protein analysis (shown below) together with *nahG* transgenic plants and *npr1* mutants showed that SA is essential for PR protein synthesis and for most of *Arabidopsis* disease resistance.

PR proteins and disease resistance

Although PR expression is a result of pathogen-host interactions, the involvement of PR proteins in disease resistance has not been conclusive. In plants transgenic for *PR* genes, most data showed a tolerance, but not complete resistance to pathogens (Cutt et al. 1989; Alexander et al. 1993). Recently, rapid development in *R* gene cloning and *R* gene signaling has given us more insight into the PR gene expression and function. The Pto-binding and -phophorylating transcription factors Pti4/5/6 can bind to the GCC box of PR protein genes and activate the PR protein expression (Zhou et al. 1997; Thara et al. 1999; Gu et at. 2000). Kinkema et al. (2000) found that nuclear localization of transcription factor NPR1 is required for PR gene expression.

Direct evidence of PR proteins function in disease resistance coming from the *cpr/npr1* double mutant analysis. The *cpr1/npr1* double mutant has no PR protein expression and is susceptible to the pathogens and thus indicates an NPR1-dependent *cpr1* in resistance and PR protein expression (Fig. 1.10; Bowling et al. 1994). The *cpr5/npr1* expresses both NPR1-dependent and NPR1-independent PR protein expression (Fig. 1.10 and 1.11; Bowling et al.

1997). The *cpr6* mutation is dominant and a gain-of-function mutant. The *cpr6/npr1* mutant retains the PR protein expression and abolishes resistance (Clarke et al. 1998). This result indicates that PR protein is not related to resistance in the mutant and there are an NPR1-dependent (Fig. 1.11) and an NPR1-independent (Fig. 1.10) pathways for PR protein synthesis. The differential expression of PR proteins was observed in the *sid1/eds5* and *sid2* mutants, where PR-1 is strongly reduced, but PR-2 and PR-5 express as in wild type (Nawrath et al. 1999). In conclusion, PR proteins can be activated in *R* resistance pathways, synthesized through three different pathways, and may not be directly involved in the disease resistance.

Disease resistance

Disease resistance could be SA dependent, SA independent or separated from HR. Thus, resistance can be from the R1, R2, R3, R4 or R5 pathways or any combination of these three (Fig. 1.13). For example the SA-independent resistance of *Cf* genes (Brading et al. 2000) could be combination of R1 or R2 combination with R3 and R5 because of involvement in HR, calcium dependent MAP kinase, and PR protein synthesis. The *Arabidopsis* SA-dependent *cpr6* resistance could be only R4 because of uncoupling of PR protein from resistance. The SA-dependent *RPS2* resistance is perhaps a combination of R2 and R5 because of PR protein synthesis, but no observed HR. The Pti4/5/6 PR protein pathway perhaps is the major PR protein synthesis pathway for SA-independent resistance. The fact that *Rar1* gene is not required for all *Mla* specificities, such as *Mla1*, indicate the R1 pathway is not essential for all resistance. We need to keep in mind that this model is an integrated signaling pathway developed by combining the data from individual genes and individual species. The next important thing to do is to test the related genes in other pathogens and in other species.

References

Aarts, N., Metz, M., Holub, E., Staskawicz, B.J., Daniels, M.J., and Parker, J.E. (1998) Different requirements for EDS1 and NDR1 by disease resistance genes define at least two *R* gene-mediated signaling pathways in *Arabidopsis*. Proc. Natl. Acad. Sci. USA. 95:10306-10311.

Alexander, D., Goodman, R.M., Gut-Rella, M., Glascock, C., Weymann, K., Friedrich, L., Maddox, D., Ahl-Goy, P., Luntz, T., Ward, E., et al. (1993) Increased tolerance to two oomycete pathogens in transgenic tobacco expressing pathogenesis-related protein 1a. Proc. Natl. Acad. Sci. USA. 90:7327-9331.

Anderson, P.A., Lawrence, G.J., Morrish, B.C., Ayliffe, M.A., Finnegan, E.J., and Ellis, J.G. (1997) Inactivation of the flax rust resistance gene *M* associated with loss of a repeated unit within the leucine-rich repeat coding region. Plant Cell 9:641-651.

Aravind, L., Dixit, V.M., and Koonin, E.V. (1999) The domains of death: evolution of the apoptosis machinery. Trends Biochem. Sci. 24:47-53.

Baker, B., Zambriski, P., Staskawicz, B., and Dinesh-Kumar, S.P. (1997) Signaling in plant-microbe interactions. Science 276:726-733.

Bendahmane, A., Kanyuka, K., and Baulcombe, D.C. (1999) The *Rx* gene from potato controls separate virus resistance and cell death responses. Plant Cell 11:781-792.

Bent, A.F., Kunkel, B.N., Dahlbeck, D., Brown, K.L., Schmidt, R., Giraudat, J., Leung, J., and Staskawicz, B.J. (1994) *RPS2* of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance genes. Science 265:1856-1860.

Bittner-Eddy, P.D., Crute, I.R., Holub, E.B., and Beynon, J.L. (2000) *RPP13* is a simple locus in *Arabidopsis thaliana* for alleles that specify downy mildew resistance to different avirulence determinants in *Peronospora parasitica*. Plant J. 21:177-188.

Blumwald, E. Aharon ,G.S., and Lam, B.C.H. (1998) Early signal transduction pathways in plant–pathogen interactions. Trends in Plant Science 3:342-346.

Bogdanove, A.J. and Martin, G.B. (2000) AvrPto-dependent Pto-interacting proteins and AvrPto-interacting proteins in tomato. Proc. Natl. Acad. Sci. USA. 97:8836-8840.

Botella, M.A., Parker, J.E., Frost, L.N., Bittner-Eddy, P.D., Beynon, J.L., Daniels, M.J., Holub, E.B., and Jones, J.D. (1998) Three genes of the *Arabidopsis RPP1* complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. Plant Cell 10:1847-1860.

Bourne, H.R., Sanders, D.A., and McCormick, F. (1991) The GTPase superfamily: conserved structure and molecular mechanism. Nature 349:117-127.

Brading, P.A., Hammond-Kosack, K.E., Parr, A., and Jones, J.D. (2000) Salicylic acid is not required for *Cf-2*- and *Cf-9*-dependent resistance of tomato to *Cladosporium fulvum*. Plant J. 23:305-318.

Bryngelsson, T. and Collinge, D. B. (1992) Biochemical and molecular analysis of the response of barley to infection by powdery mildew. Pages 459-480 in : Biotechnology in Agriculture. No. 5. Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology. P. R. Shewry, ed. C.A.B International, Wallingford, Oxon, UK.

Brommonschenkel, S.H., Frary, A., Frary, A., and Tanksley, S.D. (2000). The broad-spectrum tospovirus resistance gene *Sw-5* of tomato is a homolog of the root-knot nematode resistance gene *Mi*. Mol. Plant Microb. Inter. 13:1130-1138.

Bryan, G.T., Wu, K.S., Farrall, L., Jia, Y., Hershey, H.P., McAdams, S.A., Faulk, K.N., Donaldson, G.K., Tarchini, R., and Valent, B. (2000) tA single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. Plant Cell 12:2033-2046.

Buschges, R., Hollricher, K., Panstruga, R., Simons, G., Wolter, M., Frijters, A., van Daelen, R., van der Lee, T., Diergaarde, P., Groenendijk, J., Topsch, S., Vos, P., Salamini, F., and Schulze-Lefert, P. (1997) The barley *Mlo* gene: a novel control element of plant pathogen resistance. Cell 88:695-705.

Cai, D., Kleine, M., Kifle, S., Harloff, H.J., Sandal, N.N., Marcker, K.A., Klein-Lankhorst, R.M., Salentijn, E.M., Lange, W., Stiekema, W.J., Wyss, U., Grundler, F.M., and Jung, C.

(1997) Positional cloning of a gene for nematode resistance in sugar beet. Science 275:832-834.

Century, K.S., Shapiro, A.D., Repetti, P.P., Dahlbeck, D., Holub, E., and Staskawicz, B.J. (1997) NDR1, a pathogen-induced component required for *Arabidopsis* disease resistance. Science 278:1963-1965.

Cheng, P., Yang, Y., Heintzen, C., and Liu, Y. (2001) Coiled-coil domain-mediated FRQ-FRQ interaction is essential for its circadian clock function in *Neurospora*. EMBO J. 20:101-108.

Chern, M.S., Song, W.Y., Pi, L., Yadav, R., He, Z., Fitzgerald, H., Zhang, Y., Dong, X., and Ronald, P. (1999) Signaling in Rice disease resistance. Biology of Plant-microbe interaction 2:254-258.

Clarke, J.D., Liu, Y., Klessig, D.F., and Dong, X. (1998) Uncoupling *PR* gene expression from NPR1 and bacterial resistance: characterization of the dominant *Arabidopsis cpr6-1* mutant. Plant Cell 10:557-569.

Clarke, J.D., Volko, S.M., Ledford, H., Ausubel, F.M., Dong, X. (2000) Roles of salicylic acid, jasmonic acid, and ethylene in *cpr*-induced resistance in *Arabidopsis*. Plant Cell 12:2175-2190.

Collins, N., Drake, J., Ayliffe, M., Sun, Q., Ellis, J., Hulbert, S., and Pryor, T. (1999) Molecular characterization of the maize *Rp1-D* rust resistance haplotype and its mutants Plant Cell 11:1365-1376.

Cutt, J.R., Harpster, M.H., Dixon, D.C., Carr, J.P., Dunsmuir, P., and Klessig, D.F. (1989) Disease response to tobacco mosaic virus in transgenic tobacco plants that constitutively express the pathogenesis-related *PR1b* gene. Virology 173:89-97.

Cooley, M.B., Pathirana ,S, Wu, H.J., Kachroo, P., and Klessig, D.F. (2000) Members of the *Arabidopsis HRT/RPP8* family of resistance genes confer resistance to both viral and oomycete pathogens. Plant Cell 12:663-676.

Dinesh-Kumar, S.P., Tham, W.H., Baker, B.J. (2000) Structure-function analysis of the tobacco mosaic virus resistance gene *N*. Proc. Natl. Acad. Sci. USA. 97:14789-14794.

Devoto, A., Piffanelli, P., Nilsson, I., Wallin, E., Panstruga, R., von Heijne, G., Schulze-Lefert, P. (1999) Topology, subcellular localization, and sequence diversity of the Mlo family in plants. J. Biol. Chem. 274:34993-35004.

Dixon, M.S., Jones, D.A., Keddie, J.S., Thomas, C.M., Harrison, K., and Jones, J.D. (1996) The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. Cell 84:451-459.

Dixon, M.S., Hatzixanthis, K., Jones, D.A., Harrison, K., and Jones, J.D. (1998) The tomato *Cf-5* disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. Plant Cell 10:1915-1925.

Ellenberger,T.E., Brandl,C.J., Struhl,K. and Harrison,S.C. (1992) The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein–DNA complex. Cell 71:1223–1237.

Ellis, J.G., Lawrence, G.J., Luck, J.E., and Dodds, P.N. (1999) Identification of regions in alleles of the flax rust resistance gene *L* that determine differences in gene-for-gene specificity. Plant Cell 11:495-506.

Feys, B., and Parker, J.E. (2000) Inteplay of signaling pathways in plant disease resistance. Trends in Genetics 16:449-455.

Frye, C.A., Tang, D., and Innes, R.W. (2001) Negative regulation of defense responses in plants by a conserved MAPKK kinase. Proc. Natl. Acad. Sci. USA. 98:373-378.

Gassmann, W., Hinsch, M.E., and Staskawicz, B.J. (1999) The *Arabidopsis RPS4* bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. Plant J. 20:265-277.

Giese, H. (1981) Powdery mildew resistance genes in the Ml-a and Ml-k regions on barley chromosome 5. Hereditas 95:51-62.

Giese, H., Jørgensen, J.H., Jensen, H.P., and Jensen, J. (1981) Linkage relationships of ten powdery mildew resistance genes on barley chromosome 5. Hereditas 95:43-50.

Glazebrook, J. (1999) Genes controlling expression of defense response in *Arabidopsis*. Curr. Opin. Plant Biol. 2:280-286.

Grant, M., Brown, I., Adams, S., Knight, M., Ainslie, A., and Mansfield, J. (2000) The *RPM1* plant disease resistance gene facilitates a rapid and sustained increase in cytosolic calcium that is necessary for the oxidative burst and hypersensitive cell death. Plant J. 23:441-450.

Grant, M.R., Godiard, L., Straube, E., Ashfield, T., Lewald, J., Sattler, A., Innes, R.W., and Dangl, J.L. (1995) Structure of the *Arabidopsis RPM1* gene enabling dual specificity disease resistance. Science 269:843-846.

Gu, Y.Q., Yang, C., Thara ,V.K., Zhou, J., and Martin, G.B. (2000) *Pti4* is induced by ethylene and salicylic acid, and its product is phosphorylated by the Pto kinase. Plant Cell. 12:771-786.

Hammond-Kosack, K.E., and Jones, J.D.G. (1997) Plant disease resistance genes. Annu. Rev. Plant Physiol. Plant Mol. Biol. 48:575-607.

Halterman, D., Zhou, F., Wei, F., Wise, R.P., and Schulze-Lefert, P. (2001) The MLA6 coiled-coil, NBS-LRR protein confers AvrMla6-dependent resistance specificity to *Blumeria graminis f.* sp. *hordei* in barley and wheat. Plant J. 25:335-348.

Harding, S.A., Oh, S.H., and Roberts, D.M. (1997) Transgenic tobacco expressing a foreign calmodulin gene shows an enhanced production of active oxygen species. EMBO J. 16:1137-1144.

Heo, W.D., Lee, S.H., Kim, M.C., Kim, J.C., Chung, W.S., Chun, H.J., Lee, K.J., Park, C.Y., Park, H.C., Choi, J.Y., and Cho, M.J. (1999) Involvement of specific calmodulin isoforms in salicylic acid-independent activation of plant disease resistance responses. Proc. Natl. Acad. Sci. USA. 96:766-771.

Hirschi, K. (2001) Vacuolar H+/Ca2+ transport: who's directing the traffic? Trends in Plant Science 6:100-104.

Jia, Y., McAdams, S.A., Bryan, G.T., Hershey, H.P., and Valent, B. (2000) Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. EMBO J. 19:4004-4014.

Jahoor, A., and Fischbeck, G. (1993) Identification of new genes for mildew resistance of barley at the Mla locus in lines derived from Hordeum spontaneum.. Plant Breed. 110:116-122.

Johal, G.S., and Briggs, S.P. (1992) Reductase activity encoded by the *HM1* disease resistance gene in maize. Science 258:985-987.

Jones, D.A., Thomas, C.M., Hammond-Kosack, K.E., Balint-Kurti, P.J., and Jones. J.D. (1994) Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. Science 266:789-793.

Jørgensen, J. H. (1994) Genetics of powdery mildew resistance in barley. Crit. Rev. Plant Sci. 13:97-119.

Kinkema, M., Fan, W., and Dong, X. (2000) Nuclear localization of NPR1 is required for activation of PR gene expression. Plant Cell 12:2339-2350.

Kobe, B., and Deisenhofer, J. (1995) Proteins with leucine-rich repeats. Curr. Opin. Struct. Biol. 5:409-416.

Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science 240:1759–1764.

Lawrence, G.J., Finnegan, E.J., Ayliffe, M.A., and Ellis, J.G. (1995) The *L6* gene for flax rust resistance is related to the *Arabidopsis* bacterial resistance gene *RPS2* and the tobacco viral resistance gene *N*. Plant Cell 7:1195-1206.

Levine, A., Pennell, R.I., Alvarez, M.E., Palmer., R., and Lamb., C. (1996) Calcium-mediated apoptosis in a plant hypersensitive disease resistance response. Curr. Biol. 6:427-437.

Li, P., Nijhawan, D., Budihardjo, I., Srinivasula, S.M., Ahmad, M., Alnemri, E.S., and Wang, X. (1997) Cytochrome c and dATP-dependent formation of *Apaf-1/caspase-9* complex initiates an apoptotic protease cascade. Cell 91:479-489.

Lin J., and Weiss, A. (2001) T cell receptor signalling. J. Cell Sci. 114:243-244.

Lupas, A. (1996) Coiled coils: new structures and new functions. Trends Biochem. Sci.. 21:375–382.

Mahadevappa, M., Descenzo, R.A., and Wise, R.P. (1994) Recombination of alleles conferring specific resistance to powdery mildew at the Mla locus in barley. Genome 37:460-468.

Martin, G.B., Brommonschenkel, S.H., Chunwongse, J., Frary, A., Ganal, M.W., Spivey, R., Wu, T., Earle, E.D., and Tanksley, S.D. (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262:1432-1436.

McDowell, J.M., Dhandaydham, M., Long, T.A., Aarts, M.G., Goff, S., Holub, E.B., and Dangl, J.L. (1998) Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the *RPP8* locus of *Arabidopsis*. Plant Cell 10:1861-1874.

McKnight, S.L. (1991) Molecular zippers in gene regulation. Sci. Am. 264:54–64.

Medzhitov, R., Preston-Hurlburt, P., and Janeway, C.A.J. (1997) A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. Nature 388:394-397.

Meyers, B.C., Chin, D.B., Shen, K.A., Sivaramakrishnan, S., Lavelle, D.O., Zhang, Z., and Michelmore, R.W. (1998) The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. Plant Cell 10:1817-1832.

Meyers, B.C., Shen, K.A., Rohani, P., Gaut, B.S., Michelmore, R.W. (1998) Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. Plant Cell 10:1833-1846.

Michelmore, R.W., Paran, I. and Kesseli, R.V. 1991. Identification of markers linked to disease-resistance genes by bulked segregation analysis: a rapid method to detect markers in

specific genomic regions by using segregation populations. Proc. Natl. Acad. Sci. USA 88:9828-9832.

Milligan, S.B., Bodeau, J., Yaghoobi, J., Kaloshian, I., Zabel, P., and Williamson, V.M. (1998) The root knot nematode resistance gene *Mi* from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. Plant Cell 10:1307-1319.

Noel, L., Moores, T.L., van Der Biezen, E.A., Parniske, M., Daniels, M.J., Parker, J.E., and Jones, J.D. (1999) Pronounced intraspecific haplotype divergence at the *RPP5* complex disease resistance locus of *Arabidopsis*. Plant Cell 11:2099-2112.

Oldroyd, G.E.D., and Staskawicz, B.J. (1998) Genetically engineered broad-spectrum disease resistance in tomato. Proc. Natl. Acad. Sci. USA. 95:10300-10305.

Ori, N., Eshed, Y., Paran, I., Presting, G., Aviv, D., Tanksley, S., Zamir, D., and Fluhr, R. (1997) The *I2C* family from the wilt disease resistance locus *I2* belongs to the nucleotide binding, leucine-rich repeat superfamily of plant resistance genes. Plant Cell 9:521-532.

Parniske, M., Hammond-Kosack, K.E., Golstein, C., Thomas, C.M., Jones, D.A., Harrison, K., Wulff, B.B., and Jones, J.D. (1997) Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. Cell 91:821-832.

Parniske, M., and Jones, J.D. (1999) Recombination between diverged clusters of the tomato *Cf-9* plant disease resistance gene family. Proc. Natl. Acad. Sci. USA. 96:5850-5855.

Petersen, M., Brodersen, P., Naested, H., Andreasson, E., Lindhart, U., Johansen, B., Nielsen, H.B., Lacy., M., Austin, M.J., Parker, J.E., Sharma, S.B., Klessig, D.F., Martienssen, R., Mattsson, O., Jensen, A.B., and Mundy, J. (2000) Arabidopsis map kinase 4 negatively regulates systemic acquired resistance. Cell 103:1111-1120.

Rathjen, J.P., Chang, J.H., Staskawicz, B.J., and Michelmore, R.W. (1999) Constitutively active *Pto* induces a *Prf*-dependent hypersensitive response in the absence of *avrPto*. EMBO J. 18:3232-3240.

Rogers, E.E., and Ausubel, F.M. (1997) *Arabidopsis* enhanced disease susceptibility mutants exhibit enhanced susceptibility to several bacterial pathogens and alterations in *PR-1* gene expression. Plant Cell 9:305-316.

Romeis, T., Piedras, P., and Jones, J.D. (2000) Resistance gene-dependent activation of a calcium-dependent protein kinase in the plant defense response. Plant Cell 12:803-816.

Romeis, T., Piedras. P., Zhang, S., Klessig, D.F., Hirt, H., and Jones, J.D. (1999) Rapid Avr9- and Cf-9 -dependent activation of MAP kinases in tobacco cell cultures and leaves: convergence of resistance gene, elicitor, wound, and salicylate responses. Plant Cell 11:273-287.

Rossi, M., Goggin, F.L., Milligan, S.B., Kaloshian, I., Ullman, D.E., and Williamson, V.M. (1998) The nematode resistance gene *Mi* of tomato confers resistance against the potato aphid. Proc. Natl. Acad. Sci. USA. 95:9750-9754.

Salmeron, J.M., Oldroyd, G.E., Rommens, C.M., Scofield, S.R., Kim, H.S., Lavelle, D.T., Dahlbeck, D., and Staskawicz, B.J. (1996) Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster. Cell 86:123-133.

Saraste, M., Sibbald, P.R., and Wittinghofer, A., (1990) The P-loop--a common motif in ATP- and GTP-binding proteins. Trends Biochem. Sci. 15:430-4344

Scott, K.J. (1992) The molecular analysis of barley resistance to powdery mildew. Pages 481-496 in : Biotechnology in Agriculture. No. 5. Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology. P. R. Shewry, ed. C.A.B International, Wallingford, Oxon, UK.

Shan, L., Thara, V.K., Martin, G.B., Zhou, J.M., and Tang, X. (2000) The *Pseudomonas* AvrPto Protein is differentially recognized by tomato and tobacco and is localized to the plant plasma membrane. Plant Cell:12:2323-2338.

Song, W.Y., Wang, G.L., Chen, L.L., Kim, H.S., Pi, L.Y., Holsten, T., Gardner, J., Wang, B., Zhai, W.X., Zhu, L.H., Fauquet, C., and Ronald, P. (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 270:1804-1806.

Tai, T.H., Dahlbeck, D., Clark, E.T., Gajiwala, P., Pasion, R., Whalen, M.C., Stall, R.E., and Staskawicz, B.J. (1999) Expression of the *Bs2* pepper gene confers resistance to bacterial spot disease in tomato. Proc. Natl. Acad. Sci. USA. 96:14153-14158.

Takken, F.L., Schipper, D., Nijkamp, H.J., and Hille, J. (1998) Identification and Ds-tagged isolation of a new gene at the *Cf-4* locus of tomato involved in disease resistance to *Cladosporium fulvum* race 5. Plant J. 14:401-411.

Takken, F.L., Thomas, C.M., Joosten, M.H., Golstein, C., Westerink, N., Hille, J., Nijkamp, H.J., De Wit, P.J., and Jones, J.D. (1999) A second gene at the tomato *Cf-4* locus confers resistance to *Cladosporium fulvum* through recognition of a novel avirulence determinant. Plant J. 20:279-288.

Tang, X., Xie, M., Kim, Y.J., Zhou, J., Klessig, D.F., Martin, G.B. (1999) Overexpression of *Pto* activates defense responses and confers broad resistance. Plant Cell 11:15-29.

Tao, Y., Yuan, F., Leister, R.T., Ausubel, F.M., Katagiri, F. (2000) Mutational analysis of the *Arabidopsis* nucleotide binding site-leucine-rich repeat resistance gene *RPS2*. Plant Cell 12:2541-2554.

Thara, V.K., Tang, X., Gu, Y.Q., Martin, G.B., and Zhou, J.M. (1999) *Pseudomonas syringae* pv *tomato* induces the expression of tomato EREBP-like genes *Pti4* and *Pti5* independent of ethylene, salicylate and jasmonate. Plant J. 20:475-483.

Van der Biezen, E.A., and Jones, J.D.G. (1998) The NB-ARC domain: A novel signaling motif shared by plant resistance gene products and regulators of cell death in animals. Curr. Biol. 8:R226-R227.

Van der Vossen, E.A. , van der Voort, J.N., Kanyuka, K., Bendahmane, A., Sandbrink, H., and Baulcombe, D.C., Bakker, J., Stiekema, W.J., and Klein-Lankhorst, R.M. (2000) Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: a virus and a nematode. Plant J. 23:567-576.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T. van de, Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407-4414

Vos, P., Simons, G., Jesse, T., Wijbrandi, J., Heinen, L., Hogers, R., Frijters, A., Groenendijk, J., Diergaarde, P., Reijans, M., Fierens-Onstenk, J., de Both, M., Peleman, J., Liharska, T., Hontelez, J., and Zabeau, M. (1998) The tomato *Mi-1* gene confers resistance to both root-knot nematodes and potato aphids. Nat. Biotechnol. 16:1365-1369.

Wang, G.L., Ruan, D.L., Song, W.Y., Sideris, S., Chen, L., Pi, L.Y., Zhang, S., Zhang, Z., Fauquet, C., Gaut, B.S., Whalen, M.C., and Ronald, P.C. (1998) *Xa21D* encodes a receptor-like molecule with a leucine-rich repeat domain that determines race-specific recognition and is subject to adaptive evolution. Plant Cell 10:765-779.

Wang, Z.X., Yano, M., Yamanouchi, U., Iwamoto, M., Monna, L., Hayasaka, H., Katayose, Y., and Sasaki, T. (1999) The *Pib* gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. Plant J. 19:55-64.

Warren, R.F., Henk, A., Mowery, P., Holub, E., and Innes, R.W. (1998) A mutation within the leucine-rich repeat domain of the *Arabidopsis* disease resistance gene *RPS5* partially suppresses multiple bacterial and downy mildew resistance genes. Plant Cell 10:1439-1452.

Whitham, S., Dinesh-Kumar, S.P., Choi, D., Hehl, R., Corr, C., and Baker, B. (1994) The product of the tobacco mosaic virus resistance gene *N*: similarity to toll and the interleukin-1 receptor. Cell 78:1101-1115.

Xiao, S., Ellwood, S., Calis, O., Patrick, E., Li, T., Coleman, M., and Turner, J.G. (2001) Broad-spectrum mildew resistance in *Arabidopsis thaliana* mediated by *RPW8*. Science 291:118-120.

Wise, R. P. And Ellingboe, A.H. (1985) Fine structure and instability of the Mla locus in barley. Genetics 111:113-130.

Yang, R.B., Mark, M.R., Gran, A., Huang, A., Xie, M.H., Zhang, M., Goddard, A., Wood, W.I., Gurney, A.L., and Godowski, P.J. (1998) Toll-like receptor-2 mediates lipopolysaccharide-induced cellular signalling. Nature 395:284-288.

Yoshimura, S., Yamanouchi, U., Katayose, Y., Toki, S., Wang, Z.X., Kono, I., Kurata, N., Yano, M., Iwata, N., and Sasaki, T. (1998) Expression of *Xa1*, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. Proc. Natl. Acad. Sci. USA. 95:1663-1668.

Zhang, S., and Klessig, D.F. (1998) Resistance gene *N*-mediated de novo synthesis and activation of a tobacco mitogen-activated protein kinase by tobacco mosaic virus infection. Proc. Natl. Acad. Sci. USA. 95:7433-7438.

Zhou, F., Kurth, J., Wei, F., Elliott, C., Vale, G., Yahiaoui, N., Keller, B., Somerville, S., Wise, R. P., and Schulze-lefert, P. (2001) Cell-autonomous Expression of Barley *Mla1* Confers Race-specific Resistance to the Powdery Mildew Fungus via a *Rar1* Independent Signaling Pathway. Plant Cell 13:337-350.

Zhou, J., Loh, Y.T., Bressan, R.A., and Martin, G.B. (1995) The tomato gene *Pti1* encodes a serine/threonine kinase that is phosphorylated by Pto and is involved in the hypersensitive response. Cell 83:925-935.

Zhou, J., Tang, X., and Martin, G.B. (1997) The Pto kinase conferring resistance to tomato bacterial speck disease interacts with proteins that bind a *cis*-element of pathogenesis-related genes. EMBO J. 16:3207-3218.

CHAPTER 2. THE *MLA* (POWDERY MILDEW) RESISTANCE CLUSTER IS ASSOCIATED WITH THREE NBS-LRR GENE FAMILIES AND SUPPRESSED RECOMBINATION WITHIN A 240-KB DNA INTERVAL ON CHROMOSOME 5S (1HS) OF BARLEY

A paper published in Genetics[1]

Fusheng Wei[2,3,4], Karin Gobelman-Werner[3,5], Shaun M. Morroll[3,5], Joachim Kurth[6], Long Mao[7], Rod Wing[7], Dario Leister[6], Paul Schulze-Lefert[6], and Roger P. Wise[2,3,5,8]

Abstract

Powdery mildew of barley, caused by *Erysiphe graminis* f. sp. *hordei*, is a model system for investigating the mechanism of gene-for-gene interaction between large-genome cereals and obligate-fungal pathogens. A large number of loci that confer resistance to this disease

---

[1] Reprinted with permission of Genetics, 1999, 153 (12), 1929-1948.

[2] Interdepartmental Genetics Program; [3]Department of Plant Pathology, Iowa State University, Ames, IA 50011-1020

[4] Primary research and author

[5] Corn Insects and Crop Genetics Research, USDA-ARS, Iowa State University, Ames, IA 50011-1020

[6] Sainsbury Laboratory, John Innes Centre, Colney Lane, Norwich NR4 7UH, U.K.

[7] Clemson University Genomics Institute, Clemson, SC 29634

[8] Author for correspondence

are located on the short arm of chromosome 5(1H). The *Mla* resistance-gene cluster is positioned near the telomeric end of this chromosome arm. AFLP-, RAPD-, and RFLP-derived markers were used to saturate the *Mla* region in a high-resolution recombinant population segregating for the (*Mla6* + *Mla14*) and (*Mla13* + *Ml-Ru3*) resistance specificities. These tightly linked genetic markers were used to identify and develop a physical contig of YAC and BAC clones spanning the *Mla* cluster. Three distinct NBS-LRR resistance-gene homologue (*RGH*) families were revealed via computational analysis of low-pass and BAC-end sequence data derived from *Mla*-spanning clones. Genetic and physical mapping delimited the *Mla*-associated, NBS-LRR gene families to a 240-kb interval. Recombination within the *RGH* families was at least ten-fold less frequent than between markers directly adjacent to the *Mla* cluster.

Introduction

Genes in plants that confer resistance to fungal pathogens frequently display characteristic gene-for-gene specificity, as originally described by FLOR (1956). In nature, there are many resistance *(R)* genes in the host, each with unique specificities to particular pathogen isolates. These *R* genes are often tightly linked or represented by many alleles. The specificities among host-resistance determinants and their corresponding pathogen isolates have been useful for the genetic analyses of several resistance-gene clusters (SHEPHERD and MAYO 1972; PARAN *et al.* 1991; DICKINSON *et al.* 1993; JONES *et al.* 1993; SUDUPAK *et al.* 1993; KESSELI *et al.* 1994; LAWRENCE *et al.* 1995; RICHTER *et al.* 1995; HU *et al.* 1996; reviewed by ANDERSON *et al.* 1997).

A large number of *Ml* specificities, which confer resistance to the powdery mildew fungus, *Erysiphe graminis* f. sp. *hordei*, have been identified in barley, *Hordeum vulgare* L. These variants are distributed among eleven groups; *Mlat*, *Mla*, *Mlk*, *Mlnn*, *Mlra*, *MlGa*, and

Mlp on chromosome 5 (1H) (reviewed by JØRGENSEN 1994), *Mlg* and *mlo* on chromosome 4 (4H) (GÖRG *et al.* 1993; BÜSCHGES *et al.* 1997), *MlLa* on chromosome 2 (2H) (GIESE *et al.* 1993), and *Mlh* on chromosome 6 (6H) (JØRGENSEN 1994). Thirty-two specificities at the *Mla* locus have been differentiated by their specific reaction to unique isolates of *E. graminis* (GIESE 1981; GIESE *et al.* 1981; WISE and ELLINGBOE 1983, 1985; JAHOOR and FISCHBECK 1993; reviewed by JØRGENSEN 1994; KINTZIOS *et al.* 1995). Hence, due to its highly variable nature, the *Mla*-resistance cluster is an excellent model for the investigation of specific recognition in gene-for-gene interactions among small grains and obligate fungal pathogens (KEEN 1990; THOMPSON and BURDON 1992; CRUTE and PINK 1996). In our earlier studies, we developed a high-resolution recombinant population (selected from 3,600 gametes) that makes possible the simultaneous analysis of a number of specificities of the *Mla* cluster (MAHADEVAPPA *et al.* 1994). Of the thirty-two *Mla* specificities, the *Mla6*, *Mla14*, *Mla13*, and *Ml-Ru3* variants present in this recombinant population are all flanked by the *Xbcd249.1* and *Xmwg036* RFLP loci (DESCENZO *et al.* 1994; DESCENZO and WISE 1996).

In preparation for positional-cloning of the *Mla* locus, we have used random amplified polymorphic DNA (RAPD; WILLIAMS *et al.* 1990), amplified fragment length polymorphism (AFLP; VOS *et al.* 1995), restriction fragment length polymorphism (RFLP; BOTSTEIN *et al.* 1980), and sequence-tagged site (STS) methods to saturate the *Mla* region with molecular markers. We have used these markers to identify Yeast Artificial Chromosomes (YACs) from the cultivar Franka, and Bacterial Artificial Chromosomes (BACs) from the cultivar Morex, that are tightly linked to and spanning the *Mla* cluster. At least eight copies of NBS-LRR resistance-gene homologues (*RGH*) were identified from the *Mla*-spanning, Morex BAC contig. These eight *RGHs* are present in three distinct families and are dispersed throughout the 240-kb, *Mla*-spanning region.

Materials and methods

Overview of the high-resolution mapping population

The barley lines used to set up the original cross for the mapping population are nearly-isogenic, differing by one or more unique *Mla* specificities in the introgressed region (MOSEMAN 1972). Each of the lines was characterized quantitatively for their respective infection kinetics and resistance specificity (WISE and ELLINGBOE 1983). Crosses were constructed between the Franger- (C.I. 16151) and Rupee- (C.I. 16155) derived isogenic lines. C.I. 16151 contains the *Mla6* and *Mla14* specificities for resistance to *E. graminis*, whereas, C.I. 16155 contains *Mla13* and *Ml-Ru3* (JØRGENSEN 1994). The flanking endosperm storage-protein-encoding genes, *Hor1* and *Hor2*, were used to select for genetic recombinants in the *Mla* region. These polypeptides are distinctly polymorphic between the lines containing different *Mla* alleles, and recombinant phenotypes can be readily visualized by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) of endosperm tip extracts (DOLL and ANDERSEN 1981). A total of 1800 F_2 seeds (representing 3,600 F_1 gametes) were screened by this method. The final population presently consists of 286 individual F_4 homozygous lines, each representing an independent recombination event between the *Hor1* and *Hor2* loci, which are 8.1 cM apart and bracket the *Mla* cluster (MAHADEVAPPA *et al.* 1994; DESCENZO *et al.* 1994).

Powdery mildew resistance screening

Infection types (IT) were scored as described in MAHADEVAPPA *et al.* (1994). The infection types 0, 1, or 2 are considered resistant reactions while the infection types 3 or 4 are considered susceptible (WISE and ELLINGBOE 1983). The Franger- (C.I. 16151), Rupee- (C.I. 16155), Kwan- (C.I. 16143, containing *Mlk*) derived lines, in addition to Manchuria (C.I. 2330) were used as controls (MOSEMAN 1972). Families that segregated with any isolate were retested with at least 16 individuals per line. Sixteen individuals were used to ensure

99% probability of observing at least one homozygous recessive individual (MATHER 1951).

Bulk design

A 3-cM window bracketing the *Mla* cluster was defined via the recombination breakpoints in our high-resolution, recombinant population (DESCENZO *et al.* 1994). We used bulk segregant, RAPD- and AFLP-analyses (GIOVANNONI *et al.* 1991; MICHELMORE *et al.* 1991; CHURCHILL *et al.* 1993; VOS *et al.* 1995) to compare pools of 14 (for RAPD) or 16 (for AFLP) DNAs that were homogeneous within the window for either the *Mla6* and *Mla14* or the *Mla13* and *Ml-Ru3* resistance specificities.

RAPD and STS analysis

RAPD analysis was carried out using ten-base oligonucleotide primers synthesized from both Operon Technologies Inc. (Alameda, CA) and Oligonucleotide Synthesis Laboratory (University British Columbia, Vancouver, Canada). A total of forty Operon (Operon Technologies Inc.) and 699 University of British Columbia (Carlson) arbitrary nucleotide sequences were used in this analysis. Map positions of RAPD (and subsequently AFLP) polymorphisms were initially positioned via a low-resolution interval-mapping population, followed by all the recombinants between *Xbcd249.1* and *Xmwg036* in the high-resolution mapping population.

PCR amplification was performed in a 25 µl a reaction volume with a 1x reaction buffer supplied by the manufacturer [20 mM Tris-HCl (pH 8.4), 50 mM KCl], 1.5 mM $MgCl_2$, 0.001% gelatin, 0.1 mM each of dNTP, either 5 µM decamer RAPD- or 20 µM STS-primer, 50 ng of genomic DNA, and 0.625 units of *Taq* DNA polymerase (Gibco BRL, Rockville, MD). The followed programs were used for amplification; for RAPD: one cycle for 1 min at 94°; 44 cycles for 5 sec at 94°, 30 sec at 36°, 1 min at 72°; with a final extension of 9 min at 72°; for the STS analysis: one cycle for 3 min at 94°; 29 cycles for 30 sec at 94°, 1 min at

60°, 1 min at 72°; with a final extension of 4 min at 72°. All PCR amplifications were performed in a PTC-100 programmable thermocycler (MJ Research Inc., Watertown, MA). Amplification products were resolved by electrophoresis at 80 volts (V) for 4 hr on a 2% thin (3 mm) agarose gel containing 1x TBE buffer (0.089 M Tris, 0.089 M Borate, 0.002 M Na_2EDTA; SAMBROOK *et al.* 1989) and 1 µg/ml ethidium bromide.

Cloning of polymorphic RAPD fragments

DNA fragments were isolated by extracting an agarose plug with the small end of a pasteur pipet followed by placement in 100 µl sterile double distilled water (ddH_2O) to elute overnight at 4°. One µl of eluted DNA / ddH_2O solution was used as a template for re-amplification with the original 10-base oligonucleotide primer. DNA inserts were purified via a modified NA45 membrane (Schleicher & Schuell, Keene, NH) extraction, ligated into pGEM-T (Promega, Madison, WI), and transformed into the *E. coli* TB-1 host strain.

AFLP analysis

All 256 pairwise combinations of ^{33}P-labeled (NEN Life Science Products Inc., Boston, MA) *Eco*RI and *Mse*I primers (listed in Table 1) were used to screen for polymorphisms between both the pools and the parents. AFLP analysis was performed as per the AFLP instruction manual (Gibco BRL). For each pool, three µl of the pre-amplified products from each of the 16 individual lines was combined, diluted 50x, and selective amplification was carried out in the presence of ^{33}P-γ-ATP-labeled *Eco*RI primer and *Mse*I primer (as shown in Table 1). The amplified fragments were size-fractionated through a 7% acrylamide gel (Long Ranger, FMC Bioproducts, Rockland, ME) and exposed directly (without drying the gel) to Biomax XR film (Eastman Kodak, Rochester, NY) at –80° for 16-24 hr.

TABLE 1. AFLP primers used for Bulk Segregant Analysis

	Source		
		This study[b]	
Primer[a]	Gibco BRL[c]	Standard AFLP[c]	Sequence-specific AFLP
*Eco*RI series	AAC, AAG, ACA, ACC ACG, ACT, AGC, AGG	AAA, AAT, AGA, AGT ATA, ATC, ATG, ATT	GAA, GAC, GAG, GAT GTA, GTC, GTG, GTT
*Mse*I series	CAA, CAC, CAG, CAT CTA, CTC, CTG, CTT	CCA, CCC, CCG, CCT CGA, CGC, CGG, CGT	TCA, TCC, TCG, TCT TGA, TGC, TGG, TGT

[a] Designates the three-nucleotide sequence added to the 3' end of the core sequence. The core sequence of *Eco*RI series is 5'-AGACTGCGTACCAATTC-3' and the core sequence of *Mse*I series is 5'-GATGAGTCCTGAGTAA-3'.

[b] Primers other than those supplied by Gibco BRL were designed by the Wise lab and synthesized by the ISU DNA Synthesis and Sequencing Facility.

[c] Primers for both standard and sequence-specific AFLP.

Sequence-specific AFLP

The AFLP pre-amplification products were obtained through the use of the E-A / M-C, E-A / M-T, E-G / M-C, and E-G / M-T primer pairs. For selective amplification, the ^{33}P-γ-ATP labeled, long terminal repeat (LTR) sequence (5'-TGTTGGAATTATGCCCTAG-3') of the barley Bare-1- retrotransposon (WAUGH *et al.* 1997) was utilized in combination with one of the random AFLP primers (listed in Table 1).

Cloning and sequencing of AFLP DNA fragments

AFLP fragments were identified by matching the target signal on the autoradiogram with its corresponding area in the acrylamide gel. The cut gel slices were dissolved in ddH_2O overnight at 4° and the fragments were enriched using only the *Eco*RI primers via 10 cycles of PCR, followed by the amplification with both *Eco*RI and *Mse*I primers for another 30 cycles. The resulting fragment was cloned into pGEM-T cloning vector and transformed into

the *E. coli* TB1 host strain for selection of putative clones. The cloned inserts were pre-screened by direct PCR with the T7-1 (5'-AATACGACTCACTATAG-3') and SP6 (5'-GATTTAGGTGACACTATAG -3') primer pairs. Co-migration (via polyacrylamide gel electrophoresis) of the cloned AFLP insert as compared to the original genomic AFLP fragment was used for final verification. Two confirmed colonies from each cloning experiment were purified with Microcon-100 (Amicon, Bedford, MA) and sequenced using T7-2 (5'-CGACTCACTATAGGGCGAAT-3') and SP6-2 (5'-GCGTTGGGAGCTCTCCCATATGGT-3') vector primers. DNA sequencing and oligonucleotide synthesis was performed by the Iowa State University DNA Sequencing and Synthesis Facility. PCR primers were designed according to the DNA sequences of the clones with the assistance of Oligo 5.0 software (Perkin Elmer Biosystems).

Preparation of chromosomal yeast/YAC DNA

YAC clones were grown and maintained using Kiwi media (AUSUBEL *et al.* 1988). The AB1380 yeast host strain was grown and maintained using YEPD media [1% (w/v) bacto-yeast extract, 2% (w/v) bacto-peptone, 2% (w/v) glucose, 100 mg/L adenine]. Single-colony purified YAC clones were used to inoculate 25 ml of selective media and grown for 24-36 hr with shaking at 30°. Cells were harvested (5000g, 10 min at 4°) and resuspended in 5 ml 50 mM Na_2EDTA (pH 8.0). Yeast cell concentration was determined with a haemocytometer. Subsequently, cells were harvested and re-suspended at a concentration of 1 x 10^9 cells/ml in re-suspension buffer (10 mM Tris, pH 7.2, 20 mM NaCl, 50 mM Na_2EDTA). The cell suspension was pre-warmed to 50° briefly before adding Lyticase (Sigma, St. Louis, MO) to 1 mg/ml and gently mixing with pre-warmed InCert agarose (FMC) [2% (w/v) in re-suspension buffer] and removed to plug molds. Plugs were allowed to set for 10 min at 4° before being removed from the molds into 5 ml of lyticase buffer (10 mM Tris, pH 7.2, 50 mM Na_2EDTA, 1 mg/ml lyticase) per ml of plug and incubated for 1 hr at 37°. Plugs were

washed once (10 min at RT) in 1x wash buffer (20 mM Tris, pH 8.0, 50 mM Na_2EDTA) before being transfered to 5 ml of proteinase K reaction buffer (100 mM Na_2EDTA, pH 8.8, 0.2% (w/v) sodium deoxycholate, 1% (w/v) sodium lauryl sarcosine, 1 mg/ml proteinase K) per ml plug and incubated for 48 hr at 50°. Plugs were washed three times (10 ml per ml of plug; 30 min at RT) in 1x wash buffer (1 mM PMSF was included in second wash to eliminate residual proteinase K), before a final wash in 0.1x wash buffer and storage at 4°.

PFGE analysis of YAC and BAC clones

Pulsed field gel electrophoresis (PFGE) conditions were as calculated by the auto algorithm function of a Bio-Rad (Hercules, CA) Chef Mapper XA. The YAC and BAC clones were resolved by PFGE at 14° using 0.5x TBE buffer and 1% SeaKem LE agarose (FMC). PFGs were stained with ethidium bromide (1 µg/ml in ddH_2O), UV nicked (60 mJoules/cm^2), and DNA transferred to Hybond N^+ membrane (Amersham Pharmacia Biotech Inc., Piscataway, NJ) using 1.5 M NaCl, 0.4 M NaOH as the transfer buffer. YAC clones were further restriction-mapped utilizing a partial digest strategy described by BURKE *et al.* (1987) employing medium-rare and rare-cutting restriction endonucleases.

Generation of hybridization probes for the physical mapping of YACs

Sequences specific for the right and left arms of the YAC vector were amplified via PCR using pBR322 as a template. The primers used were as follows: YAC-LA-as (5'-AAGCGAGCAGGACTGGGCGG-3') in conjunction with YAC-LA-s (5'-GTCAGCGGGTGTTGGCGGGT-3') amplified a 2.5-kb product specific for the left arm of the YAC vector. YAC-RA-as (5'-CGGTTTTTTCCTGTTTGGCT-3') in conjunction with YAC-RA-s (5'-TTGTTTCGGCGTGGGTATGG-3') amplified a 1.4-kb product specific for the right arm of the YAC vector. Amplification conditions for the PTC-100 programmable

thermal cycler (M.J. Research Inc.) were as follows: one cycle for 3 min at 94°; 34 cycles for 30 sec at 94°, 30 sec at 56°, 3 min at 72°; and one final extension cycle for 10 min at 72°.

Genetic mapping of YAC-ends or other sequences derived the *Mla* region

Barley DNA was isolated from frozen tissue using a modified CTAB extraction. These DNA extractions, as well as DNA gel blot analyses, were conducted as previously described (WISE and SCHNABLE 1994). YAC ends or other sequences derived from the *Mla* region were first established as being low-copy by hybridization with strip blots of *Hin*dIII digested and resolved parental DNA. Once low-copy status had been determined, sequences were screened for RFLPs by Southern hybridization with parental DNA digested with a number of restriction enzymes to reveal which restriction endonuclease revealed a polymorphism. RFLPs were exhibited as differences between the two parental lines and were mapped by Southern hybridization with recombinants from our high-resolution mapping population.

BAC AFLP fingerprinting

BAC DNA was isolated by using the CUGI protocol [CLEMSON UNIVERSITY GENOMIC INSTITUTE (CUGI)]. The standard AFLP analysis protocol was followed except that the pre-amplification and amplification primer pairs are identical. These primers do not contain a selective base. The sequence of the *Eco*RI primer (designated E-0) is 5'-AGACTGCGTACCAATTC-3' and the sequence of the *Mse*I primer (designated M-0) is 5'-GATGAGTCCTGAGTAA-3'.

BAC-end cloning and sequencing

BAC ends were cloned via double-end rescue. Briefly, the BAC DNA (0.5 µg) was digested for 4 hr at 37° with 20 units of *Nsi*I (NEB), which does not cut within the vector, but cuts fairly frequently in the genomic-DNA insert. The reaction was inactivated for 20 min at 70°. The BAC DNAs were re-circularized by self-ligation in 200 µl overnight at 16°. The

ligated products were transformed into the *E. coli* TB1 host-strain and plated on LB-chloramphenicol plates. For sequencing, "mini-BAC" DNA was prepared according to our standard BAC purification protocol and further concentrated through Microcon-100 columns (Amicon). Sequence data was obtained with ABI Big Dye terminators (PE Biosystems) using the T7-1 or M13 reverse (designated R1; 5'-GGAAACAGCTATGACCATG-3') using 5 μg of total template DNA. The resulting sequence data were utilized for designing PCR primers for mapping and further library screening.

Low-pass BAC sequencing

Cesium chloride, density gradient purified BAC DNA was used for sequencing library construction. Two ml of BAC DNA solution [20 μg DNA, 500 μl glycerol, 200 μl 10x TM (0.5 M Tris-Cl, 150 mM $MgCl_2$)] was nebulized with N_2 at 6 psi for 2 minutes. The sheared DNA was repaired for 2 hr at 16° in a 30 μl volume. The repairing solution contained 20 μl DNA, 3 μl ligation buffer with ATP (NEB), 4 μl 2.5 μM dNTP, 1 unit of T4 DNA polymerase, 1 unit of Klenow, and 10 units of polynucleotide kinase (Boehringer Mannheim, Indianapolis, IN). The DNA fragments of 1.7-3.0-kb were recovered from a 0.8% low-melt agarose gel (Gibco BRL) and purified by 2x phenol/chloroform (3:1) extraction and ethanol precipitated. DNA fragments of 1.7-3.0-kb were agarose-gel purified and ligated to dephosphorylated *Eco*RV restricted pUC18 (Gibco-BRL) overnight at 16°. One μl of the ligation solution was transformed into 25 μl of *E. coli* electroMAX DH10B TM competent cells (Gibco BRL) by electroporation with the Cell-Porator system (Gibco-BRL; Voltage 400 V, Capacitance 330 μF, Impedance low ohms, Charge rate fast, Voltage Booster resistance 4 Kohms). White colonies were picked into 96-well microtiter plates containing LB freezing medium and ampicillin, incubated overnight at 37° and stored at -80°.

Results

Four *Mla* specificities are inseparable by recombination in the C.I. 16151 x C.I. 16155 cross

Over 30 specificities of *Mla* have been described (Reviewed by JØRGENSEN 1992; 1994 KINTZIOS *et al.* 1995). Two of these specificities are present in coupling in each of our mapping parents. The Franger-derived line, C.I. 16151, contains the *Mla6* and *Mla14* specificities, as described previously (GIESE *et al.* 1981, JØRGENSEN 1992; 1994). Likewise, the Rupee-derived line, C.I. 16155, contains the *Mla13* and *Ml-Ru3* specificities (JØRGENSEN 1992, 1994; CAFFIER *et al.* 1996). Previously, we determined the position of the *Mla6*, *Mla13*, and *Mla14* specificities (MAHADEVAPPA *et al.* 1994). The first objective of this study was to confirm the position of *Mla14* and to determine the position of the fourth specificity in our mapping population, *Ml-Ru3*.

Unique infection types (IT) in response to characterized isolates of *E. graminis* are utilized to detect the different specificities in segregating populations. In our earlier studies, we had mapped the *Mla14* specificity via inoculation of the recombinant population with isolate A27 (MAHADEVAPPA *et al.* 1994). As shown in Table 2, the C.I. 16151 line that contains *Mla14* confers an IT of 2-3n in response to isolate A27, whereas, the C.I. 16155 line containing *Mla13* imparts an IT of 0 with the same isolate. However, the 0 IT in response to *Mla13* would be predicted to be epistatic over the 2-3n IT displayed by *Mla14*. Thus, it was conceivable that some of the recombinant progeny in our segregating population would display a 0 IT in response to *Mla13*, but still contain *Mla14*. This epistasis could have complicated mapping of the *Mla14* specificity.

To map the position of *Mla14* and *Ml-Ru3*, eighty-eight homozygous families that are recombinant between *Xbcd249.1* and *Xmwg036* were utilized. Four to six progeny from each

TABLE 2. Infection type of recombinant lines in the *Xbcd249.1 - Xmwg036* interval 7 days after inoculation with *E. graminis* f. sp. *hordei*

E. graminis isolate	Host specificity[a]				Number of plants with IT				
	Mla6 C.I. 16151	*Mla14* C.I. 16151	*Mla13* C.I. 16155	*Ml-Ru3* C.I. 16155	0-1	1-2	2-3n	4	Total
A27	-	2-3n	0	-	49[b]	0	39[c]	0	88
R63	-	2-3n	4	4	0	0	39	49	88
R189	-	2-3n	-	1-2n	0	49[d]	39	0	88
5874	0	4	4	4	39[e]	0	0	49	88

[a]Infection type (IT): 0 = Immune; 1n = small necrotic flecks (0.5 mm); 1-2n = Small necrotic flecks (1 mm) with very limited sporulation; 2 = Large necrotic flecks (1.5 mm) with no sporulation; 2-3n = Large necrotic flecks (1.5 mm) with limited sporulation; 4 = Abundant sporulation. A "-" designates that this specificity cannot be detected with the designated isolate.

[b, c, d, and e] Indicates that these individuals carry *Mla13, Mla14, Ml-Ru3, and Mla6,* respectively.

homozygous recombinant family were inoculated separately with the R63, R189, A27, and 5874 isolates of *E. graminis* (shown in Table 2) and subsequently scored for IT using a 0-4 scale. Isolate R63 imparts a unique IT only in response to *Mla14*, thus, screening with this isolate would provide an unambiguous result. First, the 49 recombinant individuals which displayed an IT of 0 in response to A27 (confirming the presence of *Mla13*) also displayed an IT of 4 in response to R63, indicating that these individuals, in fact, do not contain *Mla14*. Conversely, the remaining 39 recombinant individuals that displayed an IT of 2-3n in response to A27, also displayed the identical IT in response to R63. This confirmed the presence of *Mla14* in these 39 progeny. Therefore, these additional results from the

inoculations with R63 confirmed that *Mla14* specificity cosegregates (in repulsion) with *Mla13*.

To position *Ml-Ru3*, the recombinant population was inoculated with isolate R189. The C.I. 16155 parent that contains *Ml-Ru3* displays an IT of 1-2n in response to this isolate. Forty-nine recombinant individuals displayed an IT of 1-2n in response to R189, indicating the presence of *Ml-Ru3* (Table 2). These same 49 individuals also displayed an IT of 0 in response to A27, indicating the presence of the *Mla13* specificity (in coupling). Importantly, the remaining 39 recombinant individuals that displayed an IT of 0 in response to 5874 (confirming the presence of *Mla6*), also displayed an IT of 2-3n in response to A27, R63, and R189, indicating the presence of *Mla14*. Based on these experiments, we established that the *Ml-Ru3* specificity cosegregates (in repulsion) with *Mla6* and *Mla14*. Thus, current observations indicate that all four specificities in this mapping population (of 3,600 gametes) are at the same genetic position on chromosome 5 (1H). This observation could be viewed as advantageous, because it suggested that all four of these specificities could be physically close, which would facilitate their ultimate isolation.

We also further tested two putative recombinants that were reported previously in MAHADEVAPPA *et al.* (1994). In that report, we had postulated recombination between the *Mla6* and *Mla13* specificities in two F_3-families of this same mapping cross. To review, both of these putative recombinant lines (H92S 6526 and H92S 6562 in Table 5 of MAHADEVAPPA *et al.* 1994) contained one or more recombination events between the flanking markers, *Hor1* and *Hor2*, that we were using to screen the population. In addition, these progeny families displayed an IT ratio in response to infection with isolates A27 and 5874 that was consistent with a recombination event (or gene conversion) within *Mla*. However, lack of DNA markers tightly linked to *Mla* prevented the precise fingerprinting of recombination events in our previous work. Therefore, to follow-up on our assumption, the putative recombinant

lines H92S 6526 and H92S 6562 were subjected to several progeny tests with isolates 5874 and A27. However, when these H92S 6526 and H92S 6562 progeny were genotyped with our current tightly linked markers, the intra *Hor1-Hor2* recombination events appeared to be positioned on either side of the *Mla* locus. Hence, even though the original lines repeatedly displayed non-Mendelian IT ratios, at present, we are unable to confirm our original hypothesis of recombination between *Mla6* and *Mla13* at the molecular level. It is possible that a distorted segregation of parental chromosomes caused the altered IT ratios in the F_3 in these two families.

Bulk-segregant, RAPD-analysis increases the genetic resolution flanking the *Mla* cluster

Saturation of the target interval with DNA markers is a prerequisite for physical delimitation *via* large-insert clones in the complex barley genome. A 3-cM window, defined by the *Xbcd249.1 – Xmwg036* interval (DESCENZO *et al.* 1994), was established from our high-resolution population for bulk segregant (GIOVANNONI *et al.* 1991; MICHELMORE *et al.* 1991; CHURCHILL *et al.* 1993) RAPD- and AFLP- (VOS *et al.* 1995) analyses.

Seven-hundred thirty-nine RAPD primers were used to amplify DNAs from the defined bulks. Of these, ninety-one primers produced DNA fragments that were polymorphic between C.I. 16151 and C.I. 16155, or the bulks. Eighteen recombinant lines, each possessing a unique recombination breakpoint between the *Hor1* and *Hor2* loci, were used to quickly determine if markers were positioned near the *Mla* locus. Only three of the 91 primers that produced amplified polymorphisms mapped to the region between *Hor1* and *Hor2*. Primer OPA-10 (5'-GTGATCGCAG-3') amplified a 1500-bp fragment in C.I. 16155, designated $OPA\text{-}10_{1500}$, that mapped between *Hor1* and *XChs3*. Primers UBC465 (5'-GGTCAGGGCT-3') and UBC165 (5'-GAAGGCACTG-3') amplified 950-bp and 1626-bp fragments, respectively, in pools containing the *Mla6* or *Mla14* specificities but not in pools

containing the *Mla13* or *Ml-Ru3* specificities. The 950-bp fragment was designated $UBC465_{950}$ and mapped between *XChs3* and *Xmwg068*. The 1626-bp fragment was designated $UBC165_{1626}$ and cosegregated with *Mla* in the low-resolution, interval-mapping-population described above.

Eighty-nine lines, each containing a unique recombination breakpoint in the *Xbcd249.1-Xmwg036* interval, were used to fine-map $UBC165_{1626}$ to a position 0.3 cM proximal to *Mla6*. The 1626-bp $UBC165_{1626}$-derived fragment was subsequently cloned, sequenced, and a series of PCR primers were designed. The different pair-wise combinations yielded a number of genomic-PCR products, which is likely due to the repetitive sequence represented by the 1626-bp $UBC165_{1626}$-derived fragment. The combination of primers P0 (5'-GAAGGCACTGA ATCGTTGATGG-3') and P954RC (5'-CAGTTTAGGGAAGTATTGCATC-3') produced a C.I. 16151 specific product that mapped 0.28 cM distal to *Mla*. Apparently, the primer pair P0 and P954RC uncovered a sequence-related, tightly linked copy of $UBC165_{1626}$. This amplification product consistently yielded the most stable map position and was designated *Fr1062*. The *Fr1062* PCR primers amplified the same fragment from Franka, the cultivar used in the construction of the Maltagen YAC library (KLEINE *et al.* 1993; 1997).

AFLP analysis is used to further saturate the genetic map

To further enrich the *Mla* region with markers for large-insert clone isolation, 256 AFLP primer pairs were used to screen for polymorphisms between the C.I. 16151 and C.I. 16155 mapping parents and the pools described above. Out of 22,500 AFLP fragments generated, 132 polymorphisms amplified from 104 primer pairs were observed. Seven of these polymorphic fragments mapped to the *Xbcd249.1-Xmwg036* interval on our low-resolution, interval-mapping-population. In the high-resolution analysis (Table 3), it was established

TABLE 3. AFLP markers tightly linked to the *Mla* locus

Marker designation	Primer Pair[a]		Position[b]	STS marker developed
	*Eco*RI	*Mse*I		
FW36.4	AGG	CTA	1.27 D	No
FW36.2	AGC	CTG	1.24 D	No
FW26.7	AGG	CTA	1.15 D	No
FW56.0	AAC	CTT	0.65 D	No
FW108	ACG	CAT	0.14 D	Yes
FW16.8	AAC	CAG	0.59 P	No
FW15.4	ACG	CAG	0.80 P	No

[a] The three-letter sequence designates the three selective bases added to the 3' end of the core sequence. The core sequence for *Eco*RI primers is 5'-AGACTGCGTACCAATTC-3' and the core sequence for *Mse*I primers is 5'-GATGAGTCCTGAGTAA-3'.

[b] Distance in cM distal (D) or proximal (P) to the *Mla* locus.

that the *FW108* AFLP marker was 0.14 cM distal from the *Mla* locus and allele-specific primers were developed for library screening as described below.

Development of allele-specific, AFLP-derived STS markers

First-round PCR primers were designed according to the corresponding DNA sequence of the cloned AFLP fragment. Six of the primer pairs derived from the internal sequences of the seven markers did not display a polymorphism in amplification experiments of the parental DNAs (Table 3). However, the *FW108*-derived marker displayed a potential polymorphism, amplifying a strong band in parental DNA from C.I. 16151 but a weak band in parental DNA from C.I. 16155; this same pattern was observed among the recombinants in the high-resolution mapping population.

The *FW108* PCR products from both the C.I. 16151 and C.I. 16155 parental DNAs (Figure 1A) were cloned and sequenced. Three single nucleotide polymorphisms (SNPs) were detected in the 108-bp, *FW108* fragment (Figure 2). These SNPs facilitated the design

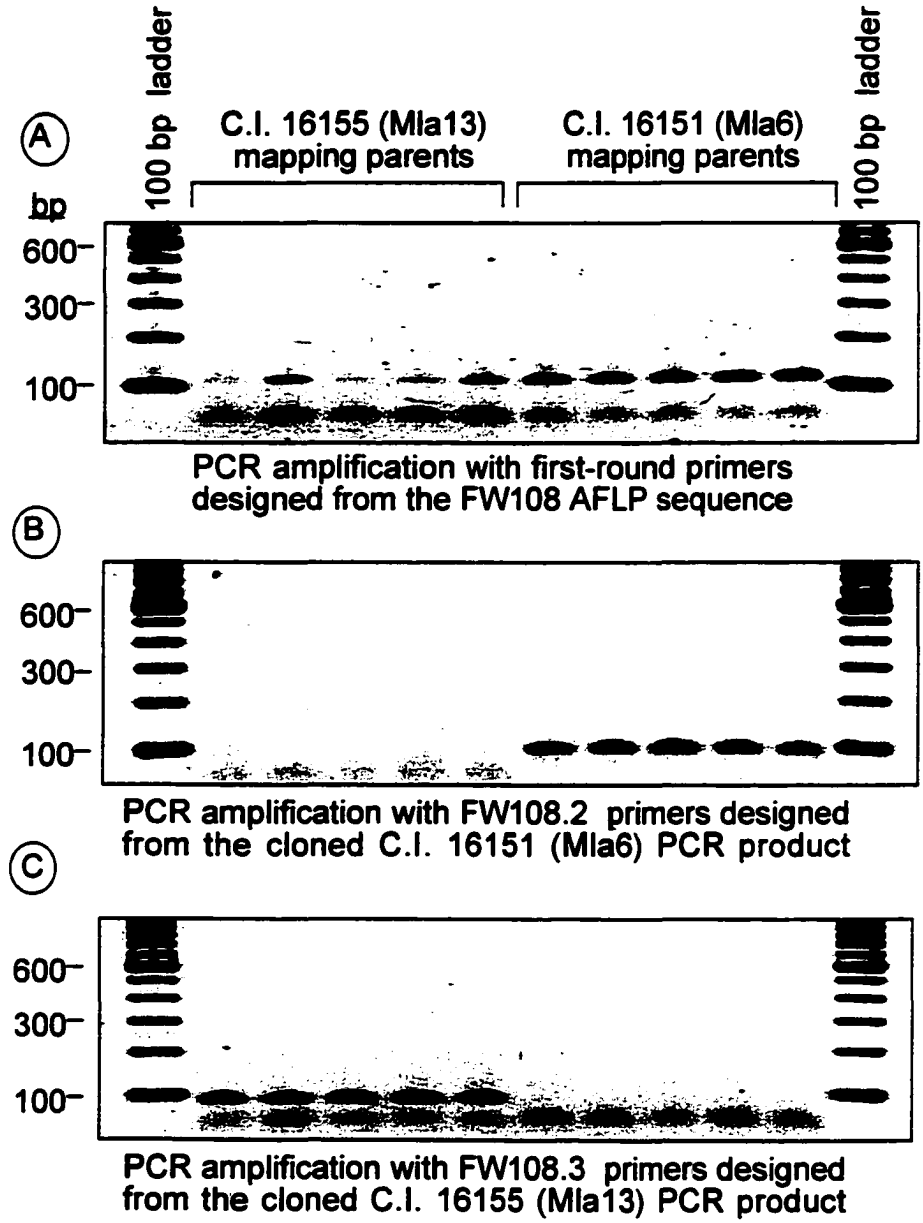


Figure. 1. PCR products amplified from C.I. 16155 and C.I. 16151 parental DNAs with *FW108* allele-specific primers. **A.** Parental-specific PCR products generated from first-round PCR primers designed according to the DNA sequence of the AFLP clone. **B.** C.I.16151-specific PCR products generated from *FW108.2* primers. **C.** C.I. 16155-specific PCR products generated from *FW108.3* primers. The *FW108.2* primers were used to screen the Maltagen (Franka) YAC library.

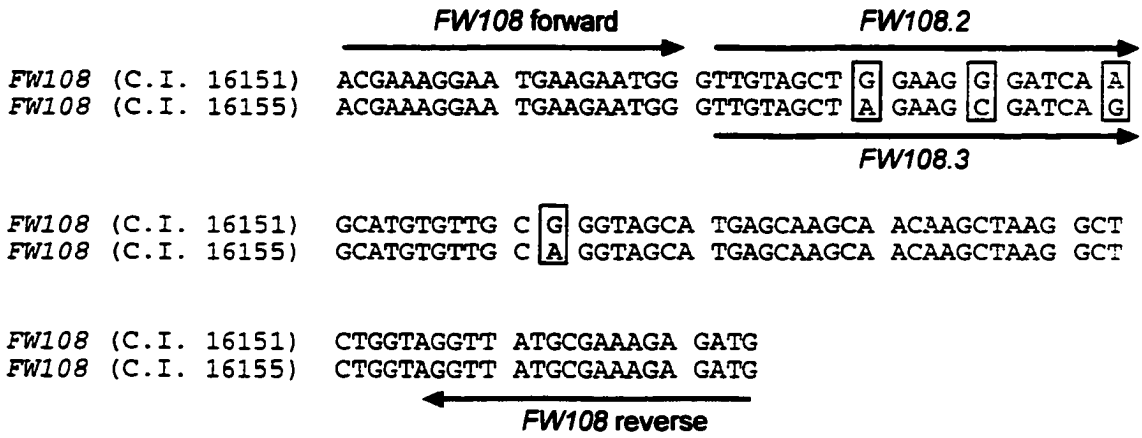


Figure. 2. Nucleotide sequence of the *FW108* fragment from C.I. 16151 and C.I. 16155. The allele-specific primers *FW108.2* and *FW108.3* were designed based upon the three-nucleotide sequence polymorphism as shown.

of C.I. 16151- and C.I. 16155-specific forward primers (designated *FW108.2* and *FW108.3*, respectively). As illustrated in Figure 1, it was established that when paired with *FW108* reverse primer, the *FW108.2* and *FW108.3* primers amplified allele-specific polymorphisms that mapped to the site of the original AFLP marker, *FW108*. This approach was not useful for the development of STS primers for the other 6 AFLP markers. The original *Eco*RI/*Mse*I polymorphism was lost when the amplified fragments were cloned and the internal sequence of the C.I. 16151 and C.I. 16155 parental fragments were 100% identical.

***BARE-1* retrotransposon, sequence-specific AFLP is used to identify additional markers**

Sequence-specific AFLP (S-SAP) was utilized to further screen for markers close to the *Mla* cluster. There are at least 3 x 10^4 copies of the *BARE-1* retrotransposon in barley, which is equivalent to 6.7% of the genome (SUONIEMI *et al.* 1996). Therefore, the *BARE-1* inverted repeat primer was used in conjunction with 24 *Eco*RI- and 24 *Mse*I-primers to amplify DNAs from the bulks and the parents. Two separate pre-amplifications were used for a total of ninety-six pairwise combinations. From the resulting 5,700 amplified bands, 114

polymorphisms were detected. One DNA fragment, designated *AgcBare*, cosegregated with the map position of *Fr1062*, 0.28 cM distal to the *Mla* locus. The number of polymorphisms detected in the bulks suggest that there are multiple, near-identical copies of the *BARE-1* retrotransposon in the *Mla*-flanking region.

The *236R* end clone from YAC236 cosegregates with the *Mla* cluster

Figure 3 illustrates the integration of all new RAPD, AFLP, and derived STS markers into the *Hor1-Mla-Hor2* region of chromosome 5 (1H). The *Fr1062* and *FW108.2* derived primers amplified DNA from Franka, and thus, markers fulfilled the criteria for large-insert clone isolation. Therefore, these two primer sets were used to screen the Maltagen (Franka) YAC library (Table 4; KLEINE *et al.* 1993; 1997). As shown in Figure 4, YAC clones were sized by PFGE followed by Southern hybridization with YAC vector-specific sequences. YAC terminal-end sequences were isolated by Inverse PCR (LEISTER *et al.* 1997b). It was established that two of these ends (*234L and 236R)* hybridized to low-copy fragments that were polymorphic between the C.I. 16151 and C.I. 16155 mapping parents. Subsequently, a combination of genetic and physical mapping established that one of the two copies of *236R* cosegregates with the *Mla* locus. During the course of this investigation, we also mapped the RFLP markers, *mwg2083* and *mwg2197,* previously shown to map between the *Hor1* and *Hor2* loci, (http://wheat.pw.usda.gov/ggpages/maps.html; kindly provided by Dr. Andreas Graner, IPK, Gatersleben, Germany). One of the three copies of *mwg2083* cosegregated with the *Mla* locus. The single-copy marker *mwg2197* was positioned two crossovers distal (0.056 cM) to the *Mla* locus. We also positioned the Resistance-Gene Homologue (*RGH*) *Hv-b6.1* (LEISTER *et al.* 1997a). We hypothesized that by taking this candidate-gene approach, we might identify large-insert clones that contained the *Mla* gene-family. This proved not to be the case as *Hv-b6.1* cosegregated with *XciwS10,* which is 0.62 cM distal to the four specificities in our mapping population (see Figure 5).

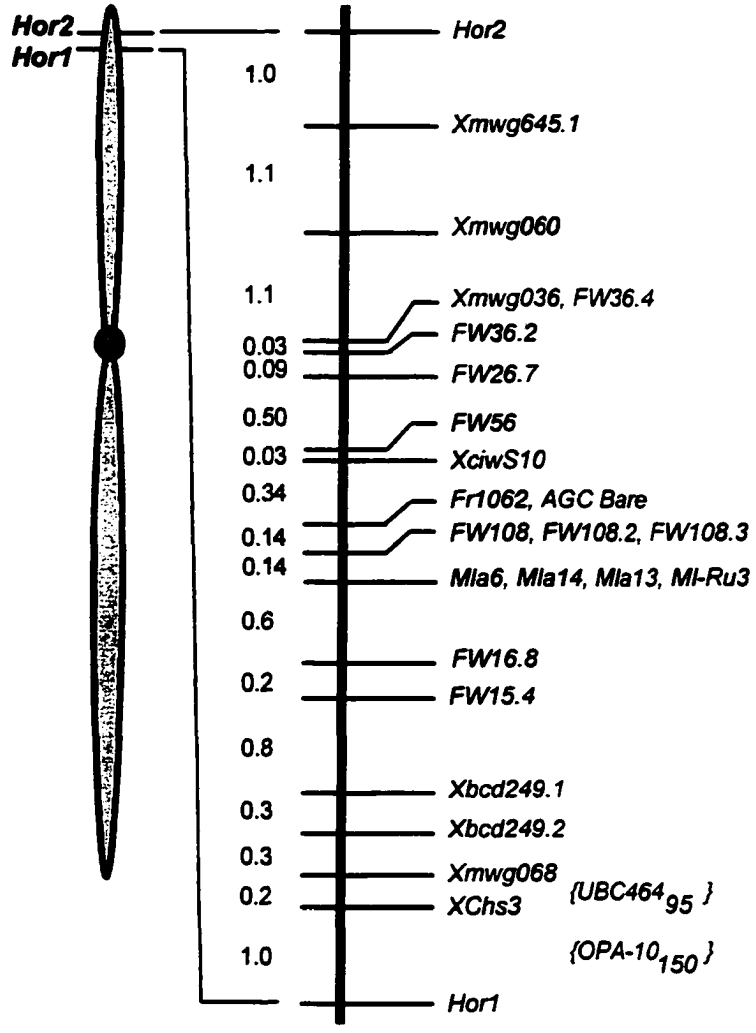


Figure. 3. High-resolution genetic map of the *Hor1-Hor2* region of barley chromosome 5 (1H). *Fr* and *FW* prefixesdesignate Wise-lab RAPD- and AFLP-derived markers, respectively. Markers in "{ }" were mapped on the low-resolution interval population. An X prefix designates an RFLP marker; *mwg* markers are from Munich-Weihenstephan-Grunbach, *bcd* markers are from Cornell University, and *ciw* markers are from the Carnegie Institute of Washington.

TABLE 4. YAC clones isolated from the Maltagen Franka library

Clone designation[a]	Identified by[b]	Insert size (kb)	Primers for right-end[c] (5' - 3')	Primers for left-end[b] (5' - 3')	Annealing temperature
75	*FW108.2*	200	AGGGCACTCTCAGGGCACTGG TCCGGATGTTCCTTGTGGTAACAC		56
101	*Fr1062*, *FW108.2*	50			
105	*FW108.2*	30	GGATCGGCGAATTCCGTATG GAATTCAAACAGATTCATAA		56
152	*FW108.2*	30/200			
234	*FW108.2*	50		TGATGTCGCGATGTTGCTCG GAATTCTGCTTGAAACCATT	56
236	*Fr1062*, *FW108.2*	280/600	CTCCTCACATCGGCTAGCTAGAT AAACCAATTAAGAGGCAGTGCAT	GAATTCTGCAGCGTTGTCCATT CCTGGCATGATCGCCTGGGCCAT	56
98IIF5	*mwg2197, 236R*	160			
99IIE7	*mwg2197, 236R*	160			
120ID1	*mwg2197, 236R*	170			

[a]YAC clones were identified by PCR amplification with *Mla* region markers.

[b]RFLP, RAPD, and AFLP markers used to identify various YACs.

[c]Primers used to amplify the ends of the designated YAC.

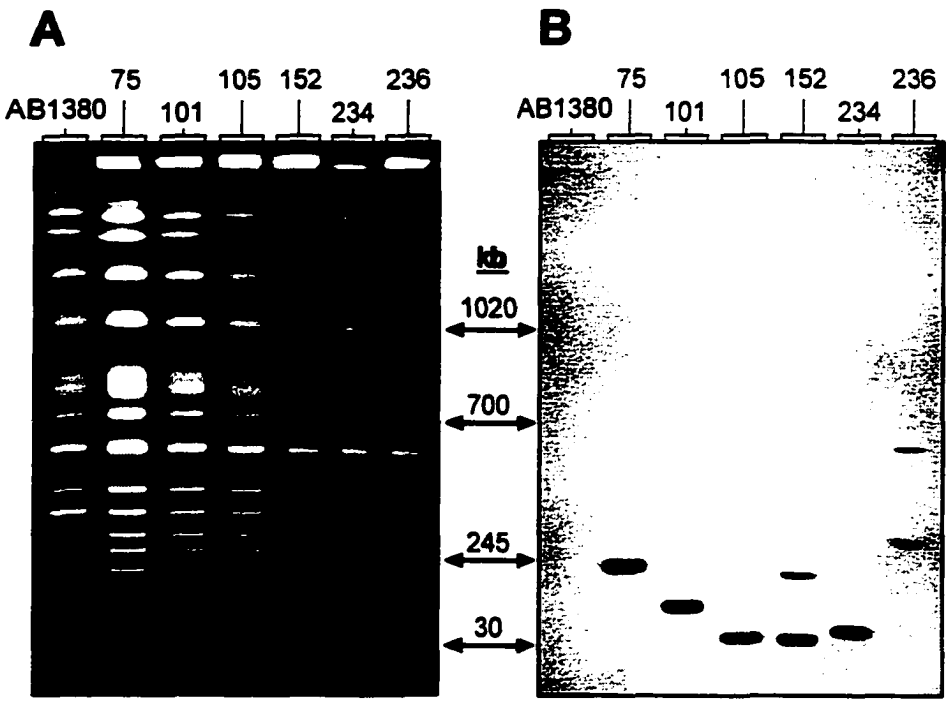
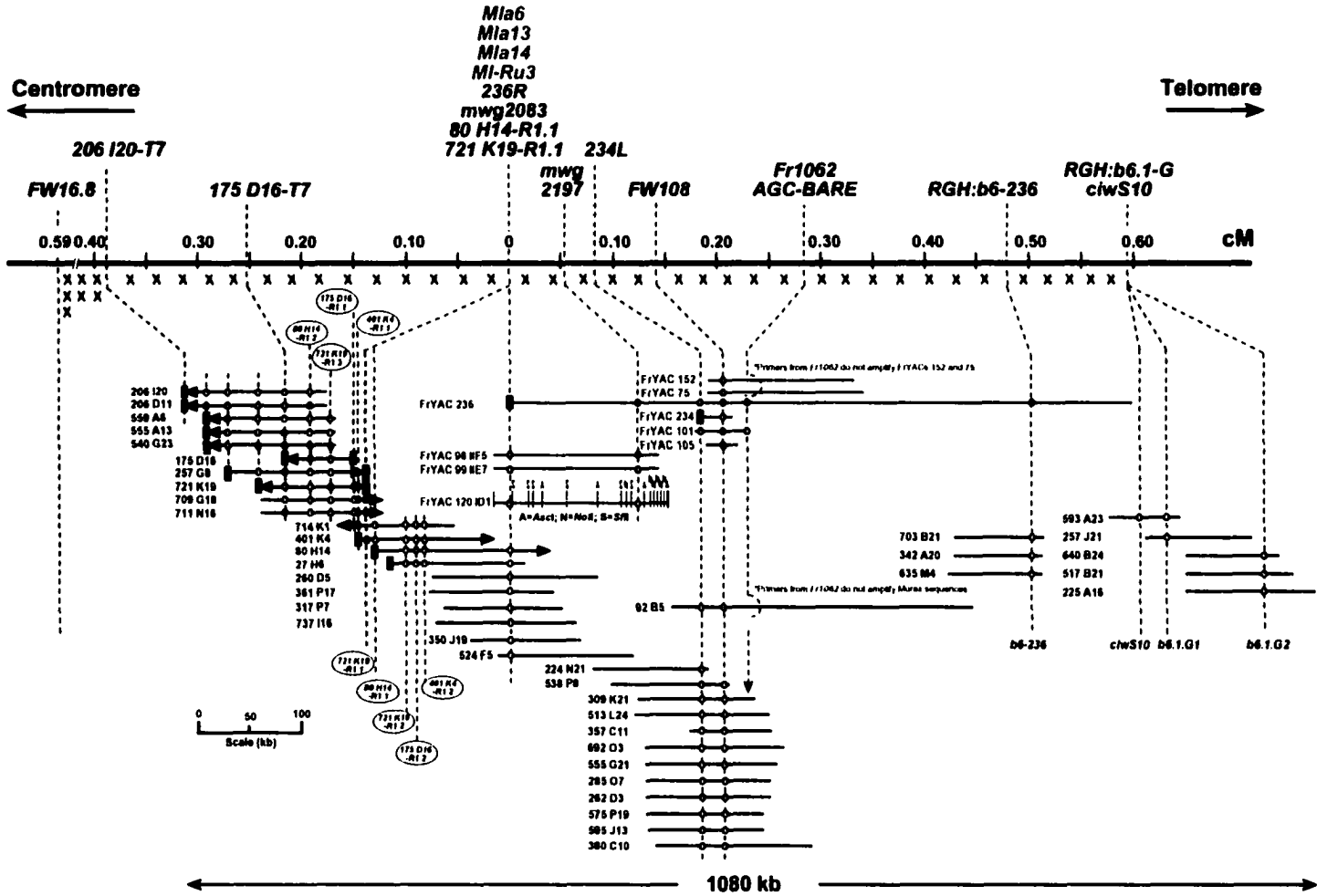


Figure 4. Analysis of Mla-linked YAC clones. High-resolution genetic analyses established that the *Fr1062*-RAPD- and *FW108*-AFLP-derived markers were 0.28 and 0.14 cM distal to the *Mla* locus, respectively, and allele-specific primers were developed to screen the Maltagen (Franka) YAC library. **A**. YAC clones were resolved by PFGE using a CHEF Mapper XA system (Bio-Rad) in conjunction with a 14-cm gel. **B**. Southern analysis of YAC clones. Filters were then hybridized with YAC vector specific sequences in order to identify all YACs.

Figure 5. Genetic and physical map of the *Mla* region. This comparison of physical to genetic distance in the *Mla* region was obtained by the use of common probes/primers on our high-resolution mapping population in addition to the overlapping Franka YACs and Morex BACs. Franka YACs are designated by a "Fr" prefix and are shown in dark red, whereas, Morex BACs are designated only by their library addresses and are shown in black. A vertical filled-in rectangle designates a cloned end-sequence from which the primers in Table 6 were developed and subsequently used for genetic and/or physical mapping. An filled-in circle designates that YAC/BAC was amplified by the respective end-clone primer set or it hybridized to the amplified product. An "X" under the top horizontal line represent crossovers in the recombinant mapping population. When BAC-ends were sequenced, horizontal arrowheads designate the T7 side of the vector. Distances are in centimorgans across the top horizontal line and YACs/BACs in the 1080-kb contig are drawn to scale in kilobases below.


Mla6
Mla13
Mla14
Ml-Ru3
236R
mwg2083
80 H14-R1.1
721 K19-R1.1
234L
Centromere
Telomere
206 I20-T7
FW16.8
175 D16-T7
mwg 2197
FW108
Fr1062 AGC-BARE
RGH:b6-236
RGH:b6.1-G ciwS10
0.59 0.40 0.30 0.20 0.10 0 0.10 0.20 0.30 0.40 0.50 0.60 cM
206 I20
206 D11
550 A6
555 A13
540 G23
175 D16
257 G8
721 K19
709 G18
711 N16
714 K1
401 K4
80 H14
27 H6
260 D5
361 P17
317 P7
737 I16
350 J19
524 F5
224 N21
538 P8
309 K21
513 L24
357 C11
692 O3
555 G21
285 O7
262 D3
575 P19
585 J13
380 C10
FrYAC 236
FrYAC 98 IIF5
FrYAC 99 IIE7
FrYAC 120 ID1
A=AscI; N=NotI; S=SfiI
FrYAC 152
FrYAC 75
FrYAC 234
FrYAC 101
FrYAC 105
92 B5
703 B21
342 A20
635 M4
593 A23
257 J21
640 B24
517 B21
225 A16
b6-236
ciwS10
b6.1.G1
b6.1.G2
0 50 100
Scale (kb)
1080 kb

Additional overlapping YACs are identified with primers developed from *236R*

For the first step in our chromosome walk to span the *Mla* cluster, primers were developed from the sequence of *236R* and *mwg2197* and used to identify the three overlapping Franka YACs, 98IIF5, 99IIE7, and 120ID1 (Table 4). YACs 98IIF5, 99IIE7, and 120ID1 were mapped physically via a partial digest strategy. This physical analysis indicated that YACs 98IIF5 and 99IIE7 contained identical DNA inserts, but were cloned in opposite orientations. YAC 120ID1 is identical to YACs 98IIF5 and 99IIE7, except for an additional ~10-kb that begins approximately 14-kb from the *mwg2197*-end of this YAC. It is likely that these three YACs correspond to YAC2197 A, B, and C reported by SCHWARZ *et al.* (1999), as they were isolated from the same library (KLEINE *et al.* 1997) with primers developed from *mwg2197*.

DNA gel blots of total YAC digests were hybridized with probes *236R*, *mwg2197* and *mwg2083* in order to physically position them on the contig. It was established that *236R* and *mwg2083* both lie within the same ~7-kb sub-region located approximately 12-kb from the left end of YAC 120ID1. Similarly, it was determined that *mwg2197* lies on a ~7-kb sub-region positioned approximately 14-kb from the right end of YAC 120ID1 (Figure 5). These restriction analyses indicated that the physical distance between *236R/mwg2083* and *mwg2197*, is approximately 120-kb. In preparation for sequencing, YAC 120ID1 was further fractionated to create a sub-genomic pBeloBAC11 library. The identified Franka BACs IV16.11, I6.24, I3.2, and VI12.7 all hybridize to *236R* and Franka BAC III12.9 hybridizes to *mwg2197*.

Overlapping Morex BACs are identified to extend the physical contig

Unfortunately, no additional YACs could be identified which allowed us to extend the Franka contig proximal to *236R*. Therefore, for the next step in our chromosome walk, amplified products from *236R*, *234L*, *mwg2083*, and *mwg2197*, were used as hybridization

probes on high-density filters of a new 6.3-genome-equivalent, BAC library from the barley cultivar Morex [Clemson University Genomic Institute (CUGI)]. The *Mla*-cosegregating markers, *236R* and *mwg2083,* each hybridized to three classes of Morex BACs. These classes most likely originated from different regions of the genome and were designated Class I (typified by 80H14, shown in Table 5), Class II (typified by 192H7, not shown), and Class III (not typified by 80H14 or 192H7, not shown). To determine which class of Morex BACs overlapped with the Franka YACs, a number of approaches were employed. First, representative members of the three classes of Morex BACs and the five YAC 120ID1-derived Franka BACs were digested with *Hin*dIII and *Eco*RI and the resulting DNAs were size fractionated via agarose-gel electrophoresis. Due to the sequence diversity between Morex and Franka, we were unable to visually determine the overlap via co-migrating *Eco*RI and *Hin*dIII restriction fragments. When a Class I or Class II Morex BAC was used as a hybridization probe, it appeared that Class I Morex BACs were more related to the Franka BACs. However, the frequency of repetitive sequences in the barley genome complicated the interpretation. Hence, we employed a BAC AFLP fingerprinting strategy to identify small, co-migrating, amplified DNA fragments. We reasoned that co-migrating amplified fragments would be sequence-related and would facilitate the identification of the overlapping region between the Franka and Morex BACs. Indeed, four co-migrating AFLPs of 275-, 281-, 595-, and 693-bp were observed among Class I (80H14-like) BACs and the YAC 120ID1-derived Franka BACs. Sequence analysis of these co-migrating DNA fragments revealed that the Class I Morex BACs were 97-98% identical to the respective sequences from the Franka BACs. Furthermore, when the low-copy 693-bp AFLP-derived fragment was used to hybridize the initial DNA gel blots described above, only Class I Morex BACs and the YAC 120ID1-derived Franka BACs showed any detectable signal. In

Table 5. Contiguous class I BAC clones isolated from the CUGI Morex library with *Mla*-spanning probes

Clone Designation	Hybridization Probe	Estimated Size (Kb)	Clone Designation	Hybridization Probe	Estimated Size (Kb)
524 F5	*236r, Mwg2083*	133	224 N21	*234l*	108
737 I16	*236r, Mwg2083*	138	538 P8	*234l*	113
317 P7	*236r, Mwg2083*	117	309 K21	*234l*	111
350 J19	*236r, Mwg2083*	105	513 L24	*234l*	133
361 P17	*236r, Mwg2083*	120	357 C11	*234l*	80
260 D5	*236r, Mwg2083*	163	692 O3	*234l*	132
27 H6	*236r, Mwg2083*	130	555 G21	*234l*	130
80 H14	*236r, Mwg2083*	165	285 O7	*234l*	120
401 K4	*80 H14-R1*	130	262 D3	*234l*	120
714 K1	*80 H14-R1*	115	575 P19	*234l*	122
711 N16	*80 H14-R1*	125	595 J13	*234l*	120
709 G18	*80 H14-R1*	125	380 C10	*234l*	152
721 K19	*80 H14-R1*	105	92 B5	*234l*	290
257 G8	*80 H14-R1*	140	342 A20	*B6-236*	83
175 D16	*80 H14-R1*	65	635 M4	*B6-236*	87
540 G23	*80 H14-R1*	122	703 B21	*B6-236*	86
555 A13	*80 H14-R1*	122	225 A16	*B6-236*	106
559 A6	*80 H14-R1*	122	517 B21	*B6-236*	92
206 D11	*80 H14-R1*	135	640 B24	*B6-236*	89
206 I20	*80 H14-R1*	135	257 J21	*B6-236*	94
			593 A23	*B6-236*	71

contrast, co-migrating AFLPs were not detected between Class II and Class III from Morex and any of the YAC 120ID1-derived BACs. These results indicated that Morex Class I BACs shown in Table 5, in fact, overlapped with the YAC 120ID1-derived Franka BACs.

Additional overlapping Morex BACs are identified that physically encompass the *Mla* cluster

For the third step in our chromosome walk, a low-copy probe developed from the 80H14-R1 end was used to identify 12 additional BACs from the Morex library. These 12 BACs all overlapped physically, due to the existence of a second copy (80H14-R1.2; see encircled markers in Figure 5) approximately 50-kb proximal to the actual R1 end of 80H14. The *Mlu*I fingerprint of these BACs shown in Figure 6 illustrates the overlapping pattern and extension of the *Mla*-spanning contig.

BAC-end sequences were used to develop primers (shown in Table 6) for genetic mapping on our high-resolution population. If an amplification polymorphism was detected with the first-round primers between our C.I. 16151 and C.I. 16155 mapping parents, then these same primers were used for mapping on every recombinant between *Xmwg036* and *Xbcd249.1* (Figure 3). This approach was utilized for the *Mla*-cosegregating STS (sequence tagged site) marker, *721K19-R1.1*. However, if no polymorphisms were observed but the first-round primers could be used to amplify a product from both the mapping parents, then the fragments were cloned and sequenced to develop allele-specific STS primers. Three additional polymorphic markers were developed by this method. As shown in Figure 5, STS marker *80H14-R1.1* co-segregated with *Mla*, STS marker *175D16-T7* was 0.25 cM proximal to *Mla*, and STS marker *206I20-T7* was 0.40 cM proximal to *Mla*.

Additionally, individual BACs were used as template for direct PCR amplification and the products were utilized in DNA gel blot hybridization to verify the overlapping pattern.

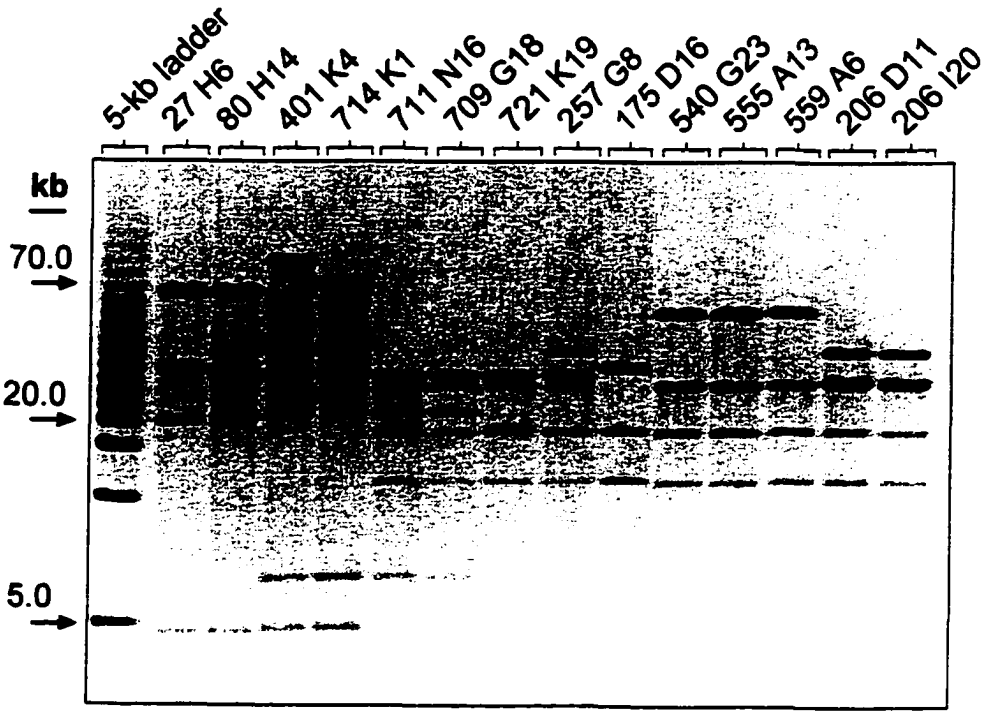


Figure 6. *Mlu* I fingerprint analysis of BAC clones. The low-copy *236R* YAC end was used to hybridize filters of the Morex BAC library. BAC 80H14 was identified as cosegregating with the *Mla* locus and a low-copy end probe developed from 80H14 was used to rescreen the library to isolate additional BAC clones. Products of digestion were resolved by PFGE using the CHEF Mapper XA system (Bio-Rad, Hercules, CA) in conjunction with a 14-cm gel.

TABLE 6. PCR primers derived from Morex BAC-end sequences and amplification on parental barley lines

BAC clones	Primers for R1 end (5' – 3')	Frag. size (bp)	Annl. temp.	Seq. origin[a]	PCR amplification on mapping parents[b] CI 16151	CI 16155	Primers for T7 end (5' – 3')	Frag. size (bp)	Seq. origin[a]	Annl. temp.	PCR amplification on mapping parents[b] CI 16151	CI 16155
80 I14	TGCTTTACCTCAAG TTGGCTGC CGAAGGTGTGTGA TTTCGATGC		54	CI 15773	+	+	GGTTACATTAGAG CACTTGCACCC CCAAGGACGAGGA AATCAGTAGG			54	+	+
	GGTGTGTGATTTC GATGCC CAGGAGCCTGCAC CGTCT	146	54	CI 16151	+	-						
	GGTGTGTGATTTC GATGCC TGCCAACTGTGTC GACGT	189	54	CI 16155	-	+						
401 K4	CAGAAGTGTTCAC TCATCCCCG AGATATAGCTTTC CATCGCACAATG	192	51	CI 15773	+	+	NA[d]	NA	NA	NA	NA	NA
711 N16 and 709G18[c]	GGAATGCAATGTA AGGCTTAAACAC GCAATCACCAGGG TACTGATATGC	310	53	CI 15773	+	+	NA	NA	NA	NA	NA	NA
721 K19	TTCAATAATCCCCT CAGTCGTAGC AACGGTTGCCAC CATCACTG	255	56	CI 15773	+	-	GGTAGTAATCAAG CCAATTCCGC GCAATCACCAGGG TACTGATATGC	457	CI 15773	53	+	+
257 G8	AATTCGTGAAGGG TTTCGTGG GGATATTCCTCA GATCGGACAAGC	447	56	CI 15773	+	+	TTCAATAATCCCCT CAGTCGTAGC AACGGTTGCCAC CATCACTG	255	CI 15773	56	+	-

TABLE 6. (continued)

BAC clones	Primers for R1 end (5' – 3')	Frag. size (bp)	Annl. temp.	Seq. origin[a]	PCR amplification on mapping parents[b] CI 16151	CI 16155	Primers for T7 end (5' – 3')	Frag. size (bp)	Seq. origin[a]	Annl. temp.	PCR amplification on mapping parents[b] CI 16151	CI 16155
175 D16	CTGACCAAAGTGA TCCTTAGCTCAG CGATAGCCATTGT GGAGTTGGAG	327	53	CI 15773	+	+	CACCAAAAACTTT TACCCTC ATACCGTGACCTC TCTGCTC	445	CI 15773	50	+	+
							TGTTTAATAATTT CAACACAAAG ATACCGTGACCTC TCTGCTC	272	CI 16151	51	+	-
540 G23 and 555 A13 and 559 A6[c]	CCATTTCAACAATC CAGTGTGCTC ACGCAAAAAACGT GGGTGC	386	70	CI 15773	-	-	CCGATGAGGGAAG CAATCTGAC AGAGCAAAAGCAG CAAAAGGC	329	CI 15773	56	+	+
206 D11 and 206 I20[c]	NA	NA	NA	NA	NA	NA	CTGGTTTGTTGTT GCTATGCGTTG TCATTTGGTGTGG GGCAAAG	470	CI 15773	56	+	+
							TTTTGATCTGATC CGGCGC TCATTTGGTGTGG GGCAAAG	298	CI 16155	56	-	+

[a] Accessions CI 16151 and CI 16155 contain the *Mla6* and *Mla13* specificities, respectively. C.I. 15773 is the designation for Morex.

[b] + designates amplification; - designates lack of amplification.

[c] BACs of identical size and inserts.

[d] NA; data not available.

All of the primers positioned between *80H14-R1.1* and *80H14-R1.2* (Figure 5) amplified an additional fragment from BAC 80H14. This result suggests that there is a large tandem duplication of 80H14 sequences on the overlapping BACs proximal to *80H14-R1.1*.

Low-pass and BAC-end sequencing reveals eight resistance-gene-homologues (*RGH*s) within a 240-kb interval

Four 96-well plates were used to sequence 384 random subclones derived from Morex BAC 80H14. These data were combined with 24 BAC-end sequences to expedite gene discovery in the *Mla*-spanning region. One-hundred and forty-four thousand nucleotides from random 80H14 subclones and 13,200 nucleotides from BAC ends were utilized for computational analyses via BLASTx searches of the NCBI non-redundant (nr) database. This approach revealed six near-full length sequences that possessed highly significant amino-acid similarity to the NBS-LRR class of cloned plant-resistance genes (RONALD 1998). Five of these sequences originated from the random sequencing of 80H14 and one was revealed by the T7-end sequencing of BAC 714K1.

Pairwise comparisons of these NBS-LRR Resistance-Gene Homologues (*RGH*) were performed using BLASTn, BLASTp (ALTSCHUL *et al.* 1997), and the GAP comparison of GCG [Wisconsin Package™ for sequence analysis (Oxford Molecular, Madison, WI)]. Initially, comparisons were delimited to sequences between and including the P-loop and the "GLPLA" motif (BAKER *et al.* 1997). Pairwise BLASTp comparisons of the deduced amino acid sequences indicate that these *RGH*s fall into three families. The *RGH1* family consists of three members, the *RGH2* family consists of two members, and the *RGH3* family has one member. Intra-family GCG-GAP comparisons revealed 80 – 98% deduced amino-acid similarity between members of *RGH* families 1 and 2, whereas inter-family comparisons showed only 46 – 51% amino-acid similarity.

We also compared the near full-length nucleotide sequences of the NBS-LRR-like *RGHs*. Inter-family BLASTn comparisons between members of *RGH* families 1, 2, and 3 revealed no significant similarity. However, the three members within the *RGH1* family are at least 60% - 98% similar and the two members within the *RGH2* family are 97% similar. Pairwise BLASTp comparisons revealed that members within a family contain 60% - 98% amino acid similarity, whereas, pairwise inter-family comparisons among members of *RGH* families 1, 2, and 3 revealed 33% amino acid similarity or less. Pairwise, intra-family GCG-GAP comparisons revealed that members within a family were up to 87% similar at both the nucleic acid and amino acid level. Pairwise, inter-family GCG-GAP similarity among members of *RGH* families 1, 2, and 3 was 47% or less at the nucleic acid level and 44% or less at the amino acid level. To the same extent, we did not observe inter-family cross-hybridization among the members of *RGH1*, *RGH2*, and *RGH3* under high-stringency wash conditions (0.1% SDS, 0.1x SSPE at 65°).

Genetic mapping and physical organization of the *RGH* families

Sequences corresponding to the *RGHs* from 80H14 were mapped back onto our high-resolution population. Allele-specific PCR primers and/or polymorphic-hybridization probes were developed for *RGH1a, 1b, 1d, 1e,* and *3a* (Table 7). We were unable to obtain allele-specific primers for *RGH1c* and *2a*, due to the monomorphic feature of the respective products amplified from the C.I. 16151 and C.I. 16155 mapping parents. Intragenic DNA sequence similarities between the C.I. 16151- and C.I. 16155-derived amplified *RGH* fragments are over 95%, whereas, the intragenic DNA sequence similarities between C.I. 16151- or C.I. 16155-derived *RGH* amplicons and Morex are about 80%. However, each of the allele-specific *RGHs* was genetically positioned in the physical interval that cosegregated with *Mla6, Mla14, Mla13*, and *Ml-Ru3*. This demonstrates that most, if not all, Morex-derived *RGHs* are also present in lines that contain characterized *Mla* specificities and that

TABLE 7. *RGH*-specific Primer Pairs

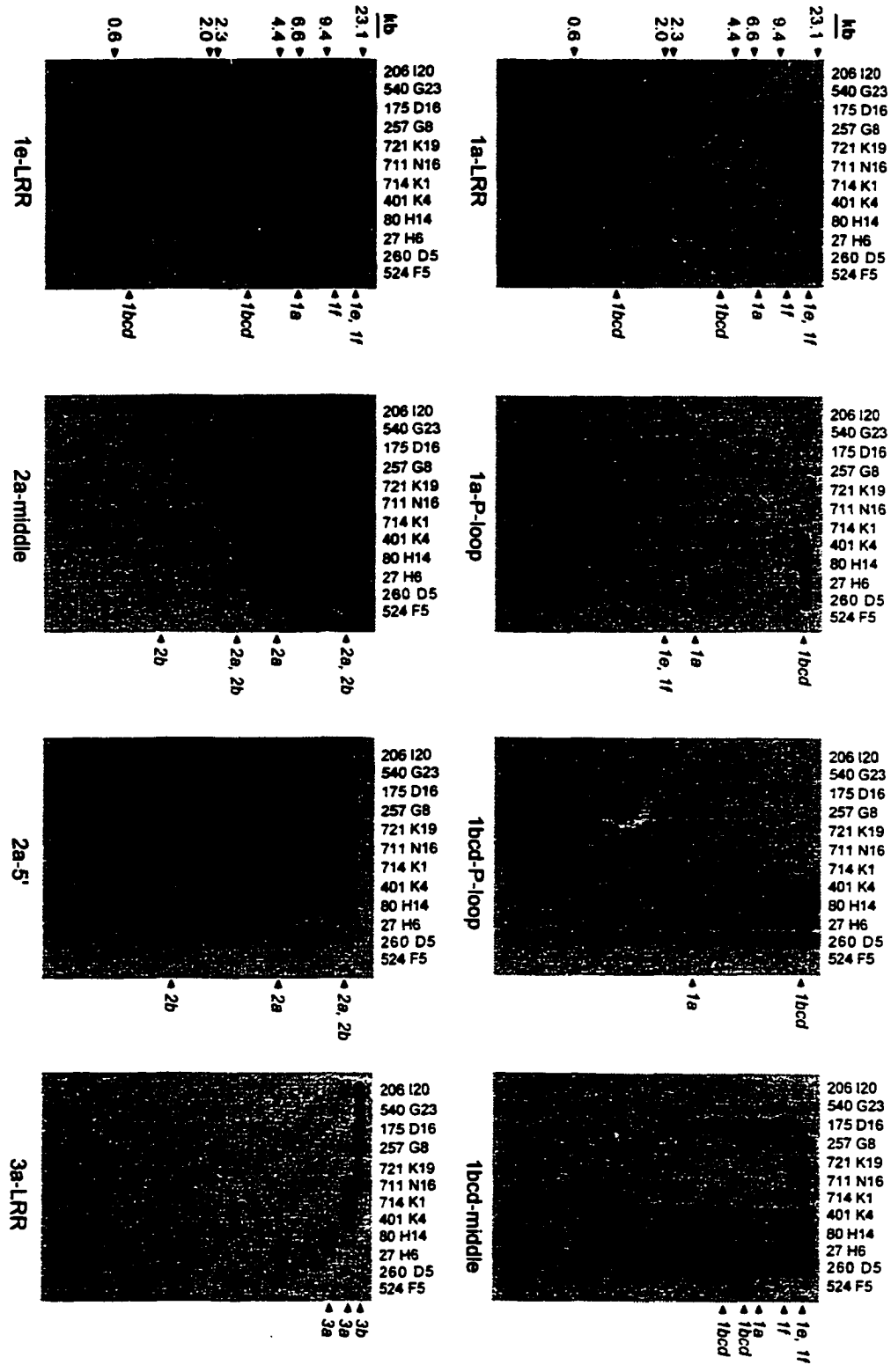
Primer designation	Primer sequences (5' -> 3')	Fragment size (bp)	*RGH* designation	Region of *RGH* ORF	Annealing temperature
39F236 39B318	TCCAATTCCGCAAGCTCTGAG ACCTCTCCTACAACATGGAGGATG	571	*RGH1a*	P-loop	56
39F146 39B184	AGCAGATACCTCGCCTAGACCATC GGTAAAGCCAAAGAGTGAATGGG	334	*RGH1a*	Middle	56
39F13 39B95	GGTTACCATCCTCTTTCGTCACC GGAGGCTCGTTGTGTCTCTGAATAC	582	*RGH1a*	LRR	56
15B02F1 15B02B9	TTTCCCAACCCACCAAATCC TCGGCGTGTTCAAGAAAGGC	307	*RGH1b*	P-loop	58
15G06F34 15G06B21	CGAGCAAAAACAGGCTAGTCCC CAACAGGAACAATTTAGGCAGTCG	527	*RGH1d*	Middle	56
38F13 38B58	GCTTGAGATCATGCACACTGTCC TGTTAGAGGCATCGTCTTCGGTC	467	*RGH1e*	P-loop	52
38F19 38B27	TGGTTCCAACTGGTGTGTTGC CCCCAATGATTTCCACGTCC	426	*RGH1e*	LRR	54
15A08F2 15A08B3	AACATGGAGATGATGGAGGCG TTGCGAGAGTATGCTGGTGAGG	750	*RGH2a*	5'	60
FWA57F FWA55B	ACCCTCGCCAGACAAGTTTACC GAAAAGCAGCTTATGCACGTCG	165	*RGH2a*	P-loop	56
38IF2 38IB4	CAGACATGCTGTGCAGATACCTCC TTCGGCTATCCACTGCTTCACC	420	*RGH2a*	Middle	54
FWA61F FWA59B	CCTTTTCTGGATTGCTCAAGTGC ATCTGGGTTGGAAGCTCCACACTC	277	*RGH2a*	LRR	56
FWA62F FWA63B	CTTTGCACTCTGGGCATAATTGAC AAGCTGGACAAGATGAAGCAGC	341	*RGH2a*	LRR	56
FWA53F FWA51B	TCGGCATCCTTATGATCCAGC TGCTCGACATCGAACCAAACC	202	*RGH3a*	P-loop	56
80H14BF3 80H14BB6	CGTGATTCCCAAGGAAATTGC GTTGCCATGTAGCTTGAGTGAGC	500	*RGH3a*	Middle	56
80H14BF30 80H14BB35	TGCTTTACCTCAAGTTGGCTGC CGAAGGTGTGTGATTTCGATGC	212	*RGH3a*	LRR	56

these *RGH* copies map to syntenic positions.

We also used sequences representing these *RGH*s as hybridization probes on BAC DNA fingerprinting filters (Figure 7), to identify additional *RGH* members on the BAC contig and derive a model for the physical organization of the *RGHs* associated with the *Mla* cluster. Indeed, another three additional *RGHs*, each belonging to one of the three families, were discovered on the adjacent BACs proximal to 80H14. By combining these hybridization results with data from the low-pass and BAC-end sequencing, at least eight *RGH*s were found in this region. These new *RGHs* fall into the three previously described families, which brings the total to four members in the *RGH1* family, two members in the *RGH2* family, and two members in the *RGH3* family. Presently, five of these NBS-LRR like *RGHs* can not be separated from the *Mla* locus by recombination events. Due to the large duplication proximal to 80H14, we were unable to develop distinct polymorphic markers between *721K19-R1.1* and *175D16-T7* (Figure 5). Therefore, at this time, we cannot determine whether the proximal three *RGHs* are genetically within the *Mla* cosegregating interval. Figure 8 illustrates a model for the physical organization of the *RGHs* associated with the *Mla* cluster. The segment between *236R* and *175D16-T7* contains all of the known *RGHs*, and therefore, defines the physical limit of *Mla*-associated, NBS-LRR gene families to 240-kb.

The use of the fingerprinting restriction endonuclease, *Eco*RI, simplified the interpretation of the physical organization of the *RGH* family members. During initial library construction, BAC inserts were ligated into the *Hin*dIII cloning site of pBeloBAC11. Restriction digestion of these BACs with the enzyme *Eco*RI releases asymmetric BAC-end fragments. Hence, migration of an *Eco*RI BAC-end fragment will be distinguishable from a BAC-internal fragment. This was advantageous because, due to the duplicated segments in the *Mla* region, additional members of a gene family could be exposed upon hybridization

Figure 7. DNA gel blot hybridization of *RGH* probes onto *Eco*RI-digested Morex BACs. All membranes were washed at high stringency (0.1x SSPE, 0.1% SDS for 30 min at 65°). Probe DNAs were amplified from the primer pairs listed in Table 7. Probes representing *RGH* domains are shown at the bottom of each panel. At the top of each panel are the BAC clone designations. At the left of panel A and E are λ-*Hin*dIII DNA size standards. *RGH* member designations are indicated on the right side of each panel. **A.** Probe 1a-LRR was amplified from primer pair 39F13 and 39B95 and corresponds to the LRR region of *RGH1a*. Four highly similar *RGH1* members hybridized to this probe. **B.** Probe 1a-P-loop was amplified from primer pair 39F236 and 39B318 and corresponds to the P-loop region of *RGH1a*. Four highly similar *RGH1* members also hybridized to this probe. **C.** Probe 1bcd-P-loop was amplified from primer pair 15B02F1 and 15B02B9 and corresponds to the P-loop region of *RGH1bcd*. The hybridization pattern revealed two highly similar members of the *RGH1* family. **D.** Probe 1bcd-middle was amplified from primer pair 15G06F34 and 15G06B21 and corresponds to the LRR region of *RGH1bcd*. The hybridization pattern indicates that there are four highly similar members of the *RGH1* family. **E.** Probe 1e-LRR was generated from primer pair 38F19 and 38B27 and corresponds to the LRR region of *RGH1e*. Four copies of the *RGH1* family hybridized to this probe. **F.** Probe 2a-middle was amplified from primer pair 38IF2 and 38IB4 and corresponds to the region between the P-loop and LRR of *RGH2a*. The hybridization result showed the existence of two highly similar members of the *RGH2* family. **G.** Probe 2a-5' was amplified from primer pair 15A08F2 and 15A08B3 and corresponds to the 5' end of the P-loop region of *RGH2a*. Two highly similar copies hybridized to this probe. **H.** Probe 3a-LRR was generated from primer pair 80H14F30 and 80H14B35 and corresponds to the LRR region of *RGH3a*. The hybridization pattern indicated that there are two copies of this region.


1a-LRR
1a-P-loop
1bcd-P-loop
1bcd-middle
1e-LRR
2a-middle
2a-5'
3a-LRR
kb
23.1
9.4
6.6
4.4
2.3
2.0
0.6
206 I20
540 G23
175 D16
257 G8
721 K19
711 N16
714 K1
401 K4
80 H14
27 H6
260 D5
524 F5
1bcd
1a
1f
1e, 1f
2b
2a
2a, 2b
3b
3a
3a, 3b

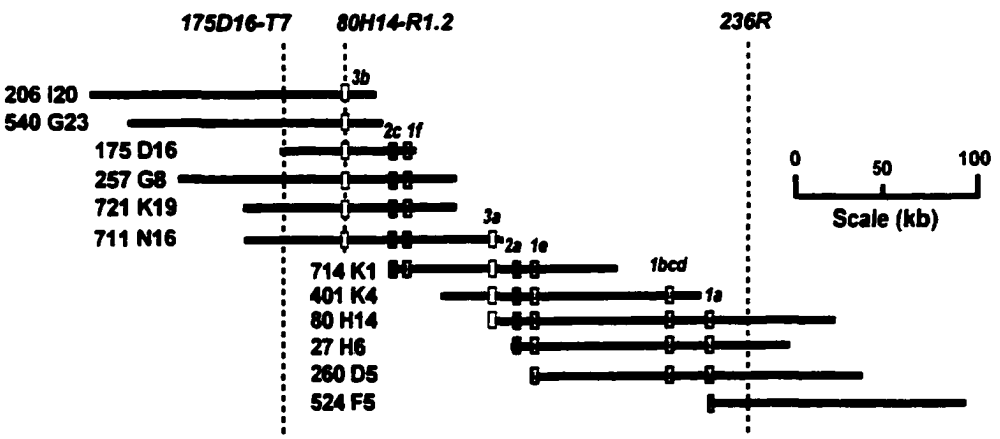


Figure 8. Physical model illustrating the minimum-tiling path of NBS-LRR resistance -gene homologues on BACs in the cosegregating *Mla* interval. This model was derived by hybridizing probes derived from the BAC-end primers listed in Table 6, the *RGH* primers listed in Table 7, and RFLP markers shown in Figure 5, to DNA-gel blots containing the EcoRI-digested BACs shown in Figure 7. There are six distinct copies of the *RGH1* family dispersed over approximately 131-kb, two copies of the *RGH2* family covering approximately 44-kb, and two copies of the *RGH3* family also covering approximately 43-kb. The two markers, *236R* and *80H14-R1.2*, define the current *Mla* cosegregating interval (Figure 5). However, the segment between *236R* and *175D16-T7* contains all of the known *RGH*s, and therefore, defines the physical limit of *Mla*-associated, NBS-LRR gene families to 240-kb.

with various *RGH* domain probes. For example, in Panels A and D of Figure 7, a *RGH1e*-comigrating fragment corresponding to *RGH1f* in BACs 711N6, 721K19, and 257G8, is revealed by the altered migration of the BAC-end from 175D16. In Panels F and G, the second member of the *RGH2* family was revealed by BAC 714K1-end sequencing and is also shown by the altered migration of the end via the *Eco*RI digest of this BAC. Furthermore, a third *RGH2* member is shown by hybridization of the non-714K1 overlapping BAC, 175D16. These additional copies were indistinguishable when DNA gel blots of *Hin*dIII restricted BACs were probed with the *RGH* probes in Figure 7.

Suppressed recombination within the *Mla* cluster

Generally, the ratio of physical to genetic to physical distance is low in regions near the centromere and high in regions towards the telomere (SCHNABLE *et al.* 1998). The average relationship between genetic and physical distance in barley, based on a genome size of 5,300 Mb and a genetic map of 1,250-1,453 cM, is 4.2-3.7 Mb/cM (KLEINHOFS *et al.* 1993, GRANER *et al.* 1991). Based on cytogenetic analysis, PEDERSEN and LINDE-LAURSEN (1995) and SOROKIN *et al.* (1994) reported a ratio of 1.0 and 2.0 Mb/cM, respectively, in the short arm region of barley chromosome 5 (1H).

As shown in Table 8, we compared physical to genetic distance ratios in eight intervals cosegregating with and adjacent to the *Mla* cluster. It appears that intervals closer to the *Mla* cluster undergo less recombination, at least in progeny of the C.I. 161561 x C.I. 16155 mapping cross. We observed no recombinants in the *236R* to *721K19-R1.1* interval, which contains nearly all the NBS-LRR, resistance-gene homologues. This lack of recombination delimits the physical to genetic distance ratio to 5 Mb/cM. However, regions immediately flanking the *Mla* cluster appeared to recombine at a higher rate than the average for the barley genome and the short arm of chromosome 5 (1H).

TABLE 8. Physical to genetic distance ratios in eight intervals spanning the *Mla* cluster

Interval	Physical distance	No. of recombinants	kb/crossover	kb/cM[a]
RGH:b6-236 -> Fr1062	~250-kb	7	36/1	1262
Fr1062 -> FW108	~25-kb	5	5/1	176
FW108 -> 234L	< 30-kb	2	15/1	< 530
234L -> mwg2197	~60-kb	1	60/1	2120
mwg2197 -> 236R (Mla)	~120-kb	2	60/1	2120
236R -> 721K19-R1.1	~140-kb	0	140/0	> 4,943
721K19-R1.1-> 175D16-T7	~80-kb	9	9/1	353
175D16-T7-> 296I20-T7	~85-kb	5	17/1	706

[a] It was previously determined that the 8.1 cM *Hor1* -> *Hor2* interval on chromosome 5S had 286 crossovers in our high resolution mapping population (DESCENZO *et al.* 1994). This is equivalent to 0.0283 cM/crossover. The kb/cM ratio in this table is based on the average resolution per recombination event in the *Hor1* to *Hor2* interval in the C.I. 16151 x C.I. 16155 cross.

Discussion

To determine the molecular processes that mediate host resistance, our aim is to isolate a number of resistance specificities of the *Mla* locus. In this report, we describe the identification of several tightly linked DNA markers and the establishment of a *Mla*-spanning, YAC and BAC contig. This *Mla*-spanning contig has facilitated the discovery of eight NBS-LRR, resistance-gene homologues, five of which cosegregate with the *Mla* locus.

Highly dissimilar NBS-LRR resistance-gene-like families are physically associated with the *Mla* cluster

In the past several years, major long-term efforts have reached fruition in the cloning of resistance genes in a variety of plant species (reviewed by RONALD 1998; MICHELMORE and MEYERS 1998). Although the isolated genes confer resistance to a diverse range of

pathogens, those involved in gene-for-gene interactions between host and pathogen share various conserved motifs. These include a serine-threonine protein-kinase domain, a leucine zipper (LZ), a Toll and interleukin-like receptor domain (TIR), a nucleotide binding site (NBS), and leucine-rich repeats (LRR). The most prevalent class of cloned plant-resistance genes contains the nucleotide-binding site combined with various lengths of a leucine-rich repeat. This NBS-LRR class is predicted to encode intracellular proteins (RONALD 1998; MICHELMORE and MEYERS 1998). The eight *RGHs* that are physically present on the Morex BAC contig belong to three distinct families of the NBS-LRR class of resistance genes. Five of these *RGHs* have been genetically delineated to the region that contains the *Mla6*, *Mla14*, *Mla13*, and *Ml-Ru3* specificities. Previous reports have shown that NBS-LRR genes can be physically juxtaposed to genes defining additional components of the resistance response. *Prf*, a NBS-LRR gene is located adjacent to *Pto*, encoding a serine-threonine kinase, both of which define essential components of race-specific resistance to bacterial speck disease in tomato (SALMERON *et al.* 1996). Among over 150 kilobases of DNA sequence surveyed, no sequences exhibiting similarities to kinases were identified in the contig spanning *Mla*. Thus, the present data suggest that the *Mla* locus contains only NBS-LRR type *RGHs*.

The majority of plant resistance genes appear to be organized as complex clusters. For example, the *Xa21* resistance gene family of rice and the *Cf-2* family of tomato are assembled as single, locally restricted clusters of homologous genes (SONG *et al.* 1995, 1997; DIXON *et al.*, 1996, 1998). The *Dm3* locus of lettuce and the *Cf4/Cf9* locus of tomato define two examples in which numerous related copies of resistance gene homologues are spread over several megabases within one chromosome (ANDERSON *et al.* 1996; SHEN *et al.* 1998; MEYERS *et al.* 1998a, 1998b; PARNISKE *et al.*, 1997). Finally, the related *L* and *M* genes of flax are located on different chromosomes (LAWRENCE *et al.* 1995; ANDERSON *et al.* 1997). In contrast, we have observed at *Mla*, an interspersed arrangement of three unrelated NBS-

LRR-like gene-families (Figure 8). Additionally, these three *Mla* cosegregating *RGH* families do not have significant similarity to the barley *Hv-b6 RGH* family, positioned 0.48 to 0.62 cM distal to the *Mla6, Mla13, Mla14,* and *Ml-Ru3* specificities (Figure 5). This multi-family organization of resistance genes and resistance-gene homologues is comparable to the recent report of mixed clusters of NBS-LRR *RGHs* of rice, each harboring at least two highly dissimilar NBS-LRR genes (LEISTER *et al.*, 1999).

RGH families and _Mla_ resistance specificities

The physical organization of the NBS-LRR-like sequences associated with the *Mla* locus was obtained from cultivar Morex, a Manchuria-type barley (KLEINHOFS *et al.* 1993). The cultivar Manchuria does not have any known *Mla* specificity and Morex also does not confer resistance to our isolates used for mapping the *Mla6, Mla13, Mla14,* and *Ml-Ru3* specificities. However, the co-segregating feature of the three *RGH* families within the genetically delimited (*Mla6, Mla14, Mla13, and Ml-Ru3*) interval indicates that they may be homologues of individual *Mla* resistance specificities. Indeed, it has been shown that susceptible cultivars or subspecies do harbor homologues of resistance genes at syntenic positions. The *Cf0* locus in susceptible *Lycopersicon esculentum* contains a homologue of the *Cf9* resistance gene that was introgressed from *Lycopersicon pimpinellifolium* (PARNISKE *et al.*, 1997). Likewise, a homologue of the *Xa1* resistance gene in the resistant cultivar IR-BB1 is present at the same locus in the susceptible near-isogenic line IR24 (YOSHIMURA *et al.*, 1998). Since *Mla6* and *Mla14* were introgressed from the wild barley *Hordeum spontaneum* nigr. (reviewed by JØRGENSEN 1994), it is conceivable that the *RGH* families at *Mla*, derived from the susceptible cultivar Morex, represent homologues of individual *Mla* resistance specificities.

There are, however, also cases in which susceptible lines lack homologues of resistance genes. For example, the *Xa21* bacterial-blight resistance locus that was introgressed from

wild rice, *Oryza longistaminata*, does not exist in cultivated rice, *Oryza sativa* (SONG *et al.* 1995; 1997). Similarly, *Rpm1* in *Arabidopsis* is present in ecotype Columbia but absent in at least six other naturally occurring accessions (GRANT *et al.*, 1995). This does not appear to be the case for the *Mla* cluster. As described above, we have been able to amplify homologous sequences corresponding to several of the *RGHs* from C.I. 16151 (containing *Mla6* + *Mla14*) and C.I. 16155 (containing *Mla13* + *Ml-Ru3*) with the Morex-derived, *RGH* primers described in Table 6. Additionally, these homologs genetically cosegregate with the *Mla6*, *Mla13*, *Mla14*, and *Ml-Ru3* specificities in our high-resolution mapping population. Taken together, these data provide the possibility that the Morex-derived *RGH* families represent homologues of single *Mla* resistance specificities.

Mutational studies uncovered two genes, *Rar1* and *Rar2*, required for *Mla* specified resistance responses (TORP and JØRGENSEN 1986; JØRGENSEN 1988; FREIALDENHOVEN *et al.* 1994; JØRGENSEN 1996). *Rar1*, located on barley chromosome 2, has been recently isolated and encodes a novel protein that is likely to function in disease resistance signaling (LAHAYE *et al.* 1998; P. SCHULZE-LEFERT, unpublished results). Mutants in *Rar1* and *Rar2* are required for the function of some but not all tested *Mla* specificities (JØRGENSEN 1988; 1996). Our finding of three unrelated *RGH* families at *Mla* could provide a simple explanation for the differential *Rar*-gene requirements if some of the *Mla* specificities are encoded by one *RGH* family and another set are encoded by a distinct *RGH* family. In this scenario, different NBS-LRR families would have the capacity to activate distinct downstream signaling components. The availability of altered-specificity mutants for *Mla1* (S. SOMERVILLE, unpublished), *Mla6* (R. P. WISE, unpublished), and *Mla12* (TORP and JØRGENSEN 1986), each exhibiting differential requirements for *Rar1* and *Rar2*, is expected to facilitate the identification of individual *Mla* resistance specificities and to provide a molecular basis to test our hypotheses.

Recombination is suppressed in highly polymorphic regions of the genome

The relationship between physical and genetic distance varies throughout the eukaryotic genome. This variation depends on many factors, including the composition of surrounding DNA sequences (PEDERSEN and LINDE-LAURSEN 1995; SCHMIDT *et al.* 1994; GILL *et al.* 1993; JIANG and GILL 1993; KOTA *et al.* 1993; LEITCH and HESLOP-HARRISON 1993; WERNER *et al.* 1992; LEITCH *et al.* 1991; SCHWARZACHER and HESLOP-HARRISON 1991; GUSTAFSON *et al.* 1990). As shown in Table 8, the ratio of physical to genetic distance varies over ten-fold in intervals adjacent to and cosegregating with the *Mla* cluster. Indeed, recombination was not observed in the cosegregating physical interval that encompasses the *Mla6, Mla14, Mla13, and Ml-Ru3* specificities and the three associated *RGH* families. This observation could be due to lack of pairing and subsequent strand exchange between homologous regions in the C.I. 16151 and C.I. 16155 parents of our mapping cross. These two accessions were originally chosen because of their high rate of hordein-polypeptide polymorphism and easily detectable differences in infection type (MAHADEVAPPA *et al.* 1994). However, it may be that suppression of recombination occurs within the *Mla* cluster because of this high rate of polymorphism. This recombination suppression contrasts to observations at the *Rp1* rust-resistance cluster in maize (COLLINS *et al.* 1999), where high rates of recombination and unequal crossover have been shown to be a source of new resistance specificities (RICHTER *et al.* 1995).

The *Mla6* allele was originally introgressed into cultivated barley from *Hordeum spontaneum*, a wild relative of *H. vulgare* (JØRGENSEN 1994). Suppressed recombination has been observed in other introgressed regions associated with disease resistance, such as the *Mi* (VAN DAELEN *et al.* 1993) and *Tm2-a* (GANAL *et al.* 1989) loci in tomato. There is a distinct difference in these two cases, however, as the *Mi* and *Tm2-a* loci are physically close to the

centromere where regions of heterochromatin were postulated to suppress recombination in this area of the chromosome.

In summary, we have established a detailed physical map of the *Mla*-spanning region and presented the physical organization of different members of *R*-gene homologues within the contig. New *Mla* mutants should allow us to determine the location of different members of this resistance-gene family and ultimately define specific regions of the gene (and therefore, protein) that are important in host-pathogen recognition. Determination of the sequence differences among mutant alleles will provide important clues in our long-range goal to understand the evolution and molecular mechanisms of host-pathogen interaction among members of the Gramineae and obligate biotrophs.

Acknowledgements

The authors thank Dr. Thomas Baum for critical review of the manuscript. This research was supported in part by USDA-NRI/CGP grant 98-35300-6169 and facilitated by the North American Barley Genome Mapping Project. D.L. was supported by a long-term EMBO postdoctoral fellowship and J.K. was supported by an EC biotechnology research grant. Research in the P.S.-L. lab is supported by a GATSBY foundation grant. Joint contribution of the Corn Insects & Crop Genetics Research Unit, USDA-Agricultural Research Service and the Iowa Agriculture and Home Economics Experiment Station. This is Journal Paper No. J-18538 of the Iowa Agricultural and Home Economics Experiment Station, Ames, IA., Project No. 3368, and supported by Hatch Act and State of Iowa funds.

References

ANDERSON, P. A., G. J. LAWRENCE, B. C. MORRISH, M. A. AYLIFFE, E. J. FINNEGAN and J. G. ELLIS, 1997 Inactivation of the flax rust resistance gene *M* associated with loss of a repeated unit within the leucine-rich repeat coding region. Plant Cell **9:** 641-651.

ANDERSON, P. A., P. A. OKUBURA, R. ARROYOGARCIA, B. C. MEYERS, and R. W. MICHELMORE, 1996 Molecular analysis of irradiation-induced and spontaneous deletion mutants at a disease resistance locus in *Lactuca sativa*. Mol.Gen.Genet. **251:** 316-325.

AUSUBEL, F. M., R. BRENT, R. E. KINGSTON, D. D. MOORE, J. G. SEIDMAN, J .A. SMITH and K. STRUHL, 1988 *Saccharomyces cerevisiae*. In Current Protocols in Molecular Biology. (Vol. 2, pp. 13.1.2). John Wiley and Sons.

ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHÄFFER, J. ZHANG, Z. ZHANG, W. MILLER, and D. J. LIPMAN, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl. Acids Res. **25**: 3389-3402.

BAKER, B., P. ZAMBRYSKI, B. STASKAWICZ, and S. P. DINESH-KUMAR, 1997 Signaling in plant–microbe interactions. Science **276:**726-733.

BOTSTEIN, D., R. L. WHITE, M. SKOLNICK and R. W. DAVIS, 1980 Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet **32:** 314-331.

BURKE, D. T., G. F. CARLE and M. V. OLSON, 1987 Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science **236:** 806-812.

BÜSCHGES, R., K. HOLLRICHER, R. PANSTRUGA, G. SIMONS, M. WOLTER, A. FRIJTERS, R. VAN DAELEN, J. GROENENDIJK, S. TOPSCH, P. VOS, F. SALAMINI and P. SCHULZE-LEFERT, 1997 The barley *Mlo* gene: a novel control element of plant pathogen resistance. Cell **88:** 695-705.

CAFFIER, V., C. DE VALLAVIEILLE-POPE, AND J. K. M. BROWN, 1996 Segregation of avirulences and genetic basis of infection types in *Erysiphe graminis* f. sp. *hordei.* Phytopathoology **86:** 1112-1121.

CHURCHILL, G. A., J. J. GIOVANNONI and S. D. TANKSLEY, 1993 Pooled-sampling makes high resolution mapping practical with DNA markers. Proc. Natl. Acad. Sci. **90:**16-20.

CLEMSON UNIVERSITY GENOMIC INSTITUTE (CUGI). Copyright © 1998 by CUGI. CUGI Bac Libraries. Rev. http://www.genome.clemson.edu/lib frame.html. (18 June 1999).

COLLINS, N., J. DRAKE, M. AYLIFFE, Q. SUN, J. ELLIS, S. HULBERT, and T. PRYOR, 1999 Molecular Characterization of the Maize *Rp1-D* Rust Resistance Haplotype and Its Mutants. Plant Cell **11:** 1365-1376.

CRUTE, I. R. and D. A. C. PINK, 1996 Genetics and utilization of pathogen resistance in plants. Plant Cell **8:** 1747-1755.

DESCENZO, R. A., R. P. WISE and M. MAHADEVAPPA, 1994 High resolution mapping of the *Hor1/Mla/Hor2* region on chromosome 5S in barley. Mol. Pl. Micr. Inter. **7:** 657-666.

DESCENZO, R. A. and R. P. WISE, 1996 Variation in the ratio of physical to genetic distance in intervals adjacent to the *Mla* locus on barley chromosome 1H. Mol. Gen. Genet. **251:** 472-482.

DICKINSON, M. J., D. A. JONES and J. D. G. JONES, 1993 Close linkage between the *Cf-2/Cf-5* and *Mi* resistance loci in tomato. Mol Pl Micr Inter **6:** 341-347.

DIXON, M. S., D. A. JONES, J. S. KEDDIE, C. M. THOMAS, K. HARRISON and J. D. G. JONES, 1996 The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. Cell **84:** 451-459.

DIXON, M. S., K. HATZIXANTHIS, D. A. JONES, K. HARRISON and J. D. G. JONES, 1998 The tomato *Cf-5* disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. Plant Cell **10:** 1915-1925.

DOLL, H., B. ANDERSEN, 1981 Preparation of barley storage protein, Hordein, for analytical sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Analytical Biochemistry **115:** 61-66.

FLOR, H. H., 1956 Host-parasite interaction in flax rust - Its genetics and other implications. Phytopathology **45:** 680-685.

FREIALDENHOVEN, A., B. SCHERAG, K. HOLLRICHER, D. B. COLLINGE, H. THORDAL-CHRISTENSEN and P. SCHULZE-LEFERT, 1994 *Nar-1* and *Nar-2*, two loci required for *Mla12*-specified race-specific resistance to powdery mildew. Plant Cell **6:** 983-994.

GANAL, M. W., N. YOUNG, and S. D. TANKSLEY, 1989 Pulsed field gel electrophoresis and physical mapping of the large DNA fragments in the *Tm-2a* region of chromosome 9 in tomato. Mol. Gen. Genet. **215:** 395-400.

GIESE, H., 1981 Powdery mildew resistance genes in the *Ml-a* and *Ml-k* regions on barley chromosome 5. Hereditas **95:** 51-62.

GIESE, H., A. G. HOLM-JENSEN, H. P. JENSEN and J. JENSEN, 1993 Localization of the Laevigatum powdery mildew resistance gene to barley chromosome 2 by the use of RFLP markers. Theor. Appl. Genet. **85:** 897-900.

GIESE, H., J. H. JØRGENSEN, H. P. JENSEN and J. JENSEN, 1981 Linkage relationships of ten powdery mildew resistance genes on barley chromosome 5. Hereditas **95:** 43-50.

GILL, K. S., B. S. GILL and T. R. ENDO, 1993 A chromosome region-specific mapping strategy reveals gene-rich telemetric ends in wheat. Chromosoma **102:** 374-381.

GIOVANNONI, J. J., R. A. WING, M. W. GANAL and S. D. TANKSLEY, 1991 Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. Nucl. Acids Res. **19:** 6553-6558.

GÖRG, R., K. HOLLRICHER and P. SCHULZE-LEFERT, 1993 Functional analysis and RFLP-mediated mapping of the *Mlg* resistance locus in barley. The Plant Journal **3:** 857-866.

GRANER, A., A. JAHOOR, J. SCHONDELMAIER, H. SIEDLER, K. PILLEN, G. FISCHBECK, G. WENZEL and R. G. HERRMANN, 1991 Construction of an RFLP linkage map of barley. Theor. Appl. Genet. **83:** 250-256

GRANT, M. R., L. GODIARD, E. STRAUBE, T. AHSFIELD, J. LEWALD, R. W. INNES, A. SATTLER and J. L. DANGL, 1995 Structure of the Arabidopsis *RPM1* gene enabling dual specificity disease resistance. Science **269:** 843-846.

GUSTAFSON, J. P., E. BUTLER and C. L. MCINTYRE, 1990 Physical mapping of a low-copy DNA sequence in rye (*Secale cereale* L.) Proc Natl. Acad. Sci USA **87:** 1899-1902

HU, B., T. E RICHTER, S. H. HULBERT and T. PRYOR, 1996 Disease lesion mimicry caused by mutations in the rust resistance gene *rp1*. Plant Cell **8:** 1367-1376.

JAHOOR, A. and G. FISCHBECK, 1993 Identification of new genes for mildew resistance of barley at the Mla locus in lines derived from *Hordeum spontaneum*. Plant Breed **110:** 116-122.

JIANG, J. and B. S. GILL, 1993 Sequential chromosome banding and *in situ* hybridization analysis. Genome **36:** 792-795

JONES, D. A., M. J. DICKINSON, P. J. BALINT-KURTI, M. S. DIXON and J. D. G. JONES, 1993 Complex resistance loci revealed in tomato by classical and RFLP mapping of the*Cf-2, Cf-4,Cf-5,* and *Cf-9* genes for resistance to *Cladosporium fulvum.* Mol Pl Micr Inter **6:** 347-357.

JØRGENSEN, J. H., 1988 Genetic analysis of barley mutants with modifications of powdery mildew resistance gene *Mla-12*. Genome **30:** 129-132.

JØRGENSEN, J. H., 1992 Multigene families of Powdery mildew resistance genes in locus *Mla* on barley chromosome 5. Plant Breeding **108:** 53-59.

JØRGENSEN, J. H., 1994 Genetics of powdery mildew resistance in barley. Critical Reviews in Plant Sciences **13:** 97-119.

JØRGENSEN, J. H., 1996 Effect of three suppressors on the expression of powdery mildew resistance genes in barley. Genome **39:** 492-498.

KEEN, N. T., 1990 Gene-for-gene complementarily in plant-pathogen interactions. Annu Rev Genet **24:** 447-463.

KESSELI, R. V., I. PARAN and R. W. MICHELMORE, 1994 Analysis of a detailed genetic linkage map of *Lactuca sativa* (lettuce) constructed from RFLP and RAPD markers. Genetics **136:** 1435-1446.

KINTZIOS, S., A. JAHOOR and G. FISCHBECK, 1995 Powdery-mildew-resistance genes *Mla29* and *Mla32* in *H. spontaneum* derived winter-barley lines. Plant Breeding **114:** 265-266.

KLEINE, M., W. MICHALEK, A. GRANER, R. G. HERRMANN and C. JUNG, 1993 Construction of a barley (*Hordeum vulgare*. L.) YAC library and isolation of a *Hor1*-specific clone. Mol. Gen. Genet. **240:** 265-272.

KLEINE, M., W. MICHALEK, T. DIEFENTHAL, H. DARGATZ and C. JUNG, 1997 Construction of a *Mlu*I-YAC library from barley (*Hordeum vulgare*. L.) and analysis of YAC insert terminal regions. Genome **40:** 896-902.

KLEINHOFS, A., A. KILIAN, M. A. SAGHAI MAROOF, R. M. BIYASHEV, P. HAYES, F. Q. CHEN, N. LAPITAN, A. FENWICK, T. K. BLAKE, V. KANAZIN, E. ANANIEV, L. DAHLEEN, D. KURDRNA, J. BOLLINGER, S. J. KNAPP, B. LIU, M. SORRELLS, M. HEUN, J. D. FRANCKOWIAK, D. HOFFMAN, R. SKADSEN and B. J. STEFFENSON, 1993 A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome. Theor. Appl. Genet. **86:** 705-712.

KOTA, R. S., B. S. GILL and T. R. ENDO, 1993 A cytogenetically based map of chromosome 1B in common wheat. Genome **36:** 548-554.

LAHAYE, T., K. SHIRASU and P. SCHULZE-LEFERT, 1998 Chromosome landing at the barley *Rar1* locus. Mol. Gen. Genet. **260:** 92-101.

LAWRENCE, G. J., E. J. FINNEGAN, M. A. AYLIFFE and J. G. ELLIS, 1995 The *L6* gene for flax rust resistance is related to the Arabidopsis bacterial resistance gene *RPS2* and the tobacco viral resistance gene *N*. Plant Cell **7:** 1195-1206.

LEISTER, D., J. KURTH, D. A. LAURIE, M. YANO, T. SASAKI, K. DEVOS, A. GRANER, and P. SCHULZE-LEFERT, 1997a Rapid reorganization of resistance gene homologues in cereal genomes. Proc. Natl. Acad. Sci. USA **95:** 370-375.

LEISTER, D., R. THOMPSON and C. GEBHARDT, 1997b Production of long IPCR-products facilitates the isolation of YAC-insert termini. Trends in Genetics **13:** Technical Tips Online [http://tto.trends.com/], T40066.

LEISTER, D., J. KURTH, D. A. LAURIE, M. YANO, T. SASAKI, A. GRANER, and P. SCHULZE-LEFERT, 1999 RFLP- and physical mapping of resistance gene homologoues in rice (*O. sativa*) and barley (H. vulgare). Theor. Appl. Genet. **98:** 509-520.

LEITCH, I. J. and J. S. HESLOP-HARRISON, 1993 Physical mapping of four sites of 5S rDNA sequences and one site of the α-amylase-2 gene in barley (*Hordeum vulgare*). Genome **36:** 517-523.

LEITCH, I .J., A. R. LEITCH and J. S. HESLOP-HARRISON, 1991 Physical mapping of plant DNA sequences by simultaneous *in situ* hybridization of two differently labeled probes. Genome **34:** 329-333.

MAHADEVAPPA, M., R. A. DESCENZO and R. P. WISE, 1994 Recombination of alleles conferring specific resistance to powdery mildew at the *Mla* locus in barley. Genome **37:** 460-468.

MANLY, K. F., 1993 A Macintosh program for storage and analysis of experimental genetic mapping data. Mammalian Genome **4:** 303-313.

MATHER, K., 1951 *The measurement of linkage in heredity*. John Wiley and Sons, Inc., New York.

MEYERS, B. C., D. B. CHIN, K., A. SHEN, S. SIVARAMAKRISHNAN, D. O. LAVELLE, Z. ZHANG and R. W. MICHELMORE, 1998a The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. Plant Cell **10:** 1817-1832.

MEYERS, B. C., K. A. SHEN, P. ROHANI, B. S. GAUT and R. W. MICHELMORE, 1998b Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. Plant Cell **10:** 1833-1846.

MICHELMORE , R. W. and B. C. MEYERS, 1998 Clusters of resistance genes evolve by divergent selection and a birth and death process. Genome Research 8: 1113-1130.

MICHELMORE, R. W., I. PARAN and R. V. KESSELI, 1991 Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl. Acad. Sci USA **88:** 9828-9832.

MOSEMAN, J. G., 1972 Isogenic barley lines for reaction to *Erysiphe graminis* f. sp. *hordei.* Crop Sci **12:** 681-682

PARAN, I., R. KESSELI and R. MICHELMORE, 1991 Identification of restriction fragment length polymorphism and random amplified polymorphic DNA markers linked to downy mildew resistance genes in lettuce, using near-isogenic lines. Genome **34:** 1021-1027.

PARNISKE, M., K. E. HAMMOND-KOSACK, C. GOLSTEIN, C. M. THOMAS, D. A. JONES, K. HARRISON, B. B. H. WULFF and J. D. G. JONES, 1997 Novel disease resistance specificity's result from sequence exchange between tandemly repeated gene at the *Cf-4/9* locus of tomato. Cell **91:** 821-832.

PEDERSEN, C. and I. LINDE-LAURSEN, 1995 The relationship between physical and genetic distance at the *Hor1* and *Hor2* loci of barley estimated by two colour fluorescent *in situ* hybridization. Theor. Appl. Genet. **91:** 941-946

RICHTER, T. E., A. J. PRYOR, J. L. BENNETZEN and S. H. HULBERT, 1995 New rust specificity's associated with recombination at the *Rp1* complex in maize. Genetics **141:** 373-381.

RONALD, P. C., 1998 Resistance Gene Evolution. Current Opinions in Biology. **1:** 294-298.

SALMERON, J. M., G. E. D. OLDROYD, C. M. T. ROMMENS, S. R. SCOFIELD, H-S KIN, B. T. LAVELLE, D. DAHLBECH and B. J. STASKAWICZ, 1996 Tomato *Prf* is a member of the leucine-rich repeat class of disease resistance genes and lies embedded within the *Pto* kinase gene cluster. Cell **86:** 123-133.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular cloning : a laboratory manual.* 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

SCHIMDT, T., T. SCHWARZACHER and J. S. HESLOP-HARRISON, 1994 Physical mapping of rRNA genes by fluorescent *in-situ* hybridization and structural analysis of 5S rRNA genes and intergenic spacer sequences in sugar beet (*Beta vulgaris*). Theor. Appl. Genet. **88:** 629-636.

SCHNABLE, P. S., A-P HSIA, and B. J. NIKOLAU, 1998 Genetic recombination in plants. Current Opinion in Plant Biology **1**: 123-129.

SCHWARZ, G., W. MICHALEK, V. MOHLER, G. WENZEL and A. JAHOOR, 1999 Chromosome landing at the *Mla* locus in barley (*Hordeum vulgare L.*) by means of high-resolution mapping with AFLP markers. Theor. Appl. Genet. **98:** 521-530.

SCHWARZACHER, T. and J. S. HESLOP-HARRISON, 1991 *In situ* hybridization to plant telomeres using synthetic oligomers. Genome **34:** 317-323

SHEN, K. A., B. C. MEYERS, N. M. ISLAM-FARIDI, D. B. CHIN, D. M. STELLY and R. W. MICHELMORE, 1998 Resistance gene candidates identified by PCR with degenerate oligonucleotide primers map to clusters of resistance genes in lettuce. Mol. Plant Microb. inter. **11**: 815-823.

SHEPHERD, K. W. and G. M. E. MAYO, 1972 Genes conferring specific plant disease resistance. Science **175:** 375-380.

SHIRASU, K., T. LAHAYE, M. TAN., F. ZHOU, C. AZEVEDO, and P. SCHULZE-LEFERT, 1999 A novel class of eukaryotic zinc-binding proteins is required for disease resistance signaling in barley and development in C. elegans. **Cell** 99:355-366.

SONG, W. Y., G. L. WANG, L. L. CHEN, H. S. KIM, L. Y. PI, T. HOLSTEN, J. GARDNER, B. WANG, W. X. ZHAI, L. H. ZHU, C. FAUQUET and P. RONALD, 1995 A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science **270:** 1804-1806.

SONG, W. Y., L. Y. PI, G. L. WANG, J. GARDNER, T. HOLSTEN and P. C. RONALD, 1997 Evolution of the rice *Xa21* disease resistance gene family. Plant Cell **9:** 1279-1287.

SOROKIN, A., F. MARTHE, A. HOUBEN, U. PICH, A. GRANER and G. KUNZE, 1994 Polymerase chain reaction mediated localization of RFLP clones to microisolated translocation chromosomes of barley. Genome **37:** 550-555.

SUDUPAK, M. A., J. L. BENNETZEN and S. H. HULBERT, 1993 Unequal exchange and meiotic instability of disease-resistance genes in the *Rp1* region of maize. Genetics **133:** 119-125.

SUONIEMI, A., K. ANAMTHAWAT-JONSSON, T. ARNA and A. H. SCHULMAN, 1996 Retrotransposon *BARE-1* is a major, dispersed component of the barley (*Hordeum valgare* L.) genome. Plant Mol. Biol. **30:** 1321-1329.

THOMPSON, J. N. and J. J. BURDON, 1992 Gene-for-gene coevolution between plants and parasites. Nature **360:** 121-135.

TORP, J. and J. H. JØRGENSON, 1986 Modification of barley powdery mildew resistance gene *Mla12* by induced mutation. Can. J. Genet. Cytol. **28:** 725-731.

VAN DAELEN, R. A. J. J., F. GERBENS, F. VAN RUSISSEN, J. AARTS, J. HONTELEEZ, and P. ZABEL, 1993 Long-range physical maps of two loci (*Aps-1* and *GP79*) flanking the root-

knot nematode resistance gene (*Mi*) near the centromere of tomato chromosome 6. Plant Mol. Biol. **23:** 185-192.

VOS, P., R. HOGERS, M. BLEEKER, M. REIJANS, T. LEE, M. VAN DE HORNES, A. FRITJERS, J. POT, J. PELEMAN, M. KUIPER and M. ZABEAU, 1995 AFLP: a new technique for DNA fingerprinting. Nucl. Acids Res. **23:** 4407-4414.

WAUGH, R., K. MCCLEAN, A. J. FLAVELL, S. R. PEARCE, A. KUMAR, B. B. T. THOMAN and W. POWEL, 1997 Genetic distribution of Bare-1-like transposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol. Gen. Genet. **253:** 687-694.

WERNER, J. E., T. R. ENDO and B. S. GILL, 1992 Toward a cytogenetically based physical map of the wheat genome. Proc Natl. Acad. Sci USA **89:** 11307-11311.

WILLIAMS, J. G. K., A. R. KUBELIK, K. J. LIVAK, J. A. RAFALSKI and S. V. TINGEY, 1990 DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl. Acids Res. **18:** 4407-4414.

WISE, R. P. and A. H. ELLINGBOE, 1983 Infection kinetics of *Erysiphe graminis* f. sp. *hordei* on barley with different alleles at the *Mla* locus. Phytopathology **73:** 1220-1222.

WISE, R. P. and A. H. ELLINGBOE, 1985 Fine structure and instability of the *Mla* locus in barley. Genetics **111:** 113-130.

WISE, R. P. and P. S. SCHNABLE, 1994 Mapping complementary genes in maize: Positioning the *rf1* and *rf2* nuclear-fertility restorer loci of Texas (T)-cytoplasm relative to RFLP and morphological markers. Theor Appl Genet **88:** 785-795.

YOSHIMURA, S., U. YAMANOUCHI, Y. KATAYOSE, S. TOKI, Z. X. WANG, I. KONO, N. KURATA, M. YANO, N. IWATA and T. SASAKI, 1998 Expression of *Xa1*, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. Proc Natl. Acad. Sci USA 95: 1663-1668.

CHAPTER 3. FUNCTIONAL CLUSTERING AND MOLECULAR EVOLUTION OF A PLANT-DEFENSE COMPLEX

A paper submitted to Nature Genetics

Fusheng Wei[1,2,3], Rod A. Wing[4] and Roger P. Wise[1,2,5,6]

Abstract and Introduction

Proteins that initiate defense against disease often share conserved signaling motifs in plants and animals(Baker et al. 1997; Hammond-Kosack and Jones 1997; van der Biezen et al. 1998). In addition to conservation of protein domains, genes with related functions are frequently clustered (Beck and Trowsdale, 2000). To assess the organization and content of a major defense gene complex in plants, we determined the complete sequence of a 261-kb BAC contig from barley cultivar (cv.) Morex that spans the *Mla* (powdery mildew) resistance cluster. The *Mla* region is comprised of thirty-two protein-encoding genes and two $tRNA^{ser}$ genes. Sixteen of the protein-coding sequences are plant-defense related; 12 of these are associated with defense against powdery mildew disease, but function in different signaling

[1] Interdepartmental Genetics Program; [2]Department of Plant Pathology, Iowa State University, Ames, IA 50011-1020

[3] Primary research and author

[4] Clemson University Genomics Institute, Clemson, SC 29634

[5] Corn Insects and Crop Genetics Research, USDA-ARS, Iowa State University, Ames, IA 50011-1020

[6] Author for correspondence

pathways. These defense-related and other protein-encoding sequences are organized as three gene-rich islands separated by two 40-kb complexes of nested transposable elements and a gene-poor region. Our results indicate that defense-related genes tend to cluster together in the barley genome and that the present *Mla* region evolved over a period greater than 7 million years through a series of duplication and inversion events in addition to nested transposable element insertion.

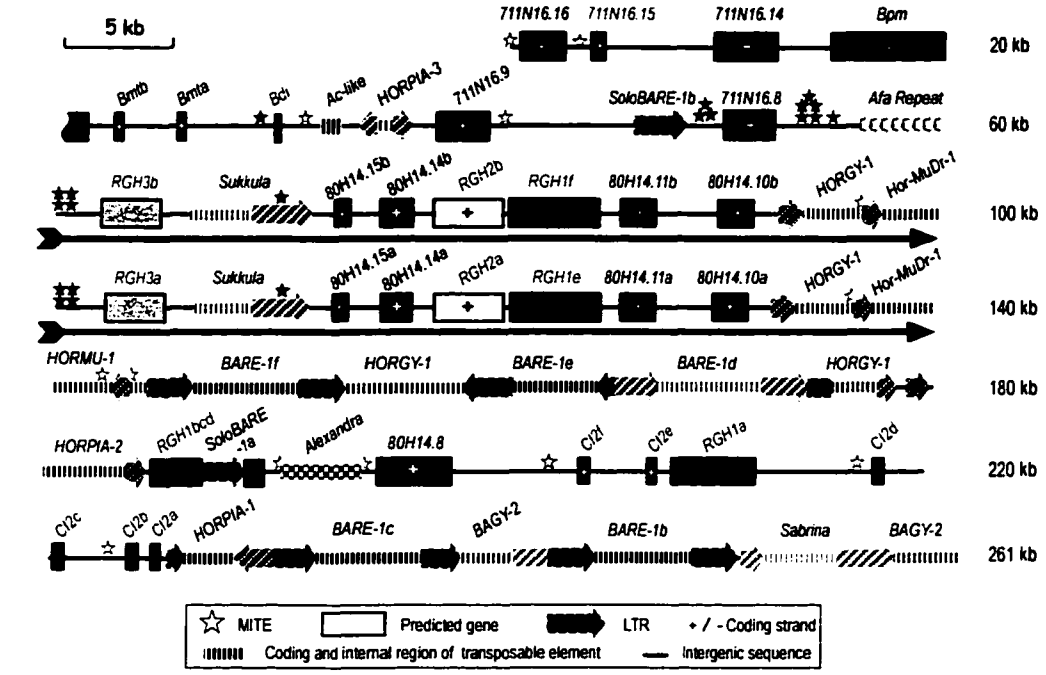
Results

Limited long-range sequence data have precluded a full understanding of how the genomes of grain cereal plants are organized, especially those with genomes that are even larger than that of humans. At the macro-genome level, recent reports have postulated the existence of gene-dense islands interrupted by gene-poor spaces (reviewed by Moore, 2000). The average gene density in the model plant genome of *Arabidopsis* is one per 4.5 kb. The genomes of grain cereal plants are orders of magnitude greater in size than *Arabidopsis*. yet would be expected to encode a similar number of genes. To assess the organization and gene content of a major *R* gene cluster in the 5000-Mb barley genome, we determined the complete 261,265 bp sequence encompassed by two overlapping BACs spanning the *Mla* (powdery mildew) resistance locus from cv. Morex.

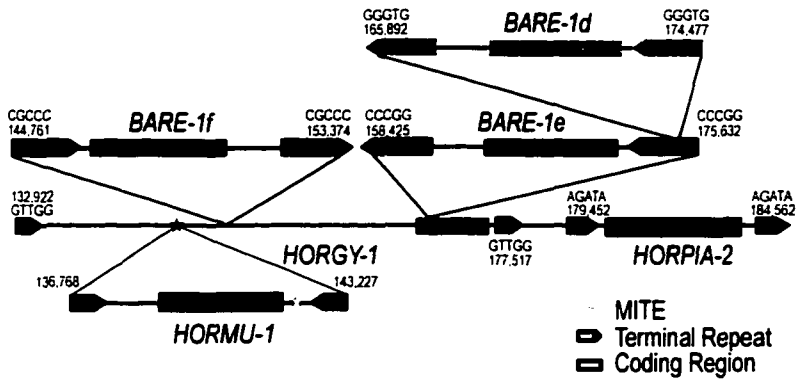
Thirty-two protein-coding and two duplicated $tRNA^{ser}$ genes were predicted by a combination of Blastx and p, GENSCAN, GeneMarker or Grail (Fig. 1a and Table 1). The overall density in the *Mla* region is one gene per 8.1-kb (Table 2). However, the density in the gene-rich islands is one gene per 4.6-kb (Table 2), statistically identical to the gene in size between the two genomes is mainly due to the composition of the intergenic space.

Fig. 1. Sequence annotation of the barley *Mla* locus from cultivar Morex. The 112,178-nt BAC 711N16 was sequenced at 15X redundancy and 158,773-nt BAC 80H14 was sequenced at 12X redundancy. A 9,686-nt overlapping sequence encompasses *RGH3a* and the 5' end of *Sukkula.1*. Different colors show different families of genes or transposable elements. **a.**) overall annotation of the 261,265-nt region. The brick text reveals defense-related genes, while the red underlined arrowhead illustrates the 40-kb tandem duplication. **b)** nested retrotransposon Complex I on BAC 80H14 between 132,922 and 184,562 nucleotides for **c)** nested retrotransposon Complex II on BAC 80H14 between 225,321 and 261,265 nucleotides. Retrotransposon placement was determined by the 5-bp LTR inverted repeat, insertion signature (6-bp direct repeat shown in letters) at the two ends of each element. This direct repeat marks the boundary for each retro element. Below the insertion signature is the position of the element.

a


5 kb
711N16.16
711N16.15
711N16.14
Bpm
20 kb
Bmtb
Bmta
Bci
Ac-like
HORPIA-3
711N16.9
SoloBARE-1b
711N16.8
Afa Repeat
60 kb
RGH3b
Sukkula
80H14.15b
80H14.14b
RGH2b
RGH1f
80H14.11b
80H14.10b
HORGY-1
Hor-MuDr-1
100 kb
RGH3a
Sukkula
80H14.15a
80H14.14a
RGH2a
RGH1e
80H14.11a
80H14.10a
HORGY-1
Hor-MuDr-1
140 kb
HORMU-1
BARE-1f
HORGY-1
BARE-1e
BARE-1d
HORGY-1
180 kb
HORPIA-2
RGH1bcd
SoloBARE-1a
Alexandra
80H14.8
Cl2f
Cl2e
RGH1a
Cl2d
220 kb
Cl2c
Cl2b
Cl2a
HORPIA-1
BARE-1c
BAGY-2
BARE-1b
Sabrina
BAGY-2
261 kb
MITE
Predicted gene
LTR
+ / - Coding strand
Coding and internal region of transposable element
Intergenic sequence

b


GGGTG
165,892
BARE-1d
GGGTG
174,477
CGCCC
144,761
BARE-1f
CGCCC
153,374
CCCGG
158,425
BARE-1e
CCCGG
175,632
132,922
GTTGG
AGATA
179,452
AGATA
184,562
HORGY-1
GTTGG
177,517
HORPIA-2
136,768
143,227
HORMU-1
MITE
Terminal Repeat
Coding Region

c

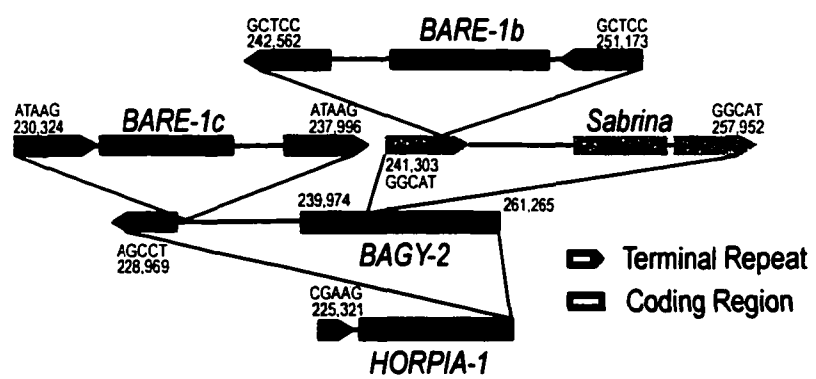

GCTCC
242,562
BARE-1b
GCTCC
251,173
ATAAG
230,324
BARE-1c
ATAAG
237,996
Sabrina
GGCAT
257,952
241,303
GGCAT
239,974
261,265
AGCCT
228,969
BAGY-2
CGAAG
225,321
HORPIA-1
Terminal Repeat
Coding Region

Table 1. Predicted Gene Products in the 261-kb, *Mla* spanning region of barley Cv. Morex

Predicted Gene Product	Position[a]		Strand	BLASTX Search			EST BLASTn search		
				Best Homolog GenBank ID	E-value	Predicted function[b]	GenBank ID	E-value	Longest Match (bp)
711N16.16	<424	2898	-	AAC17165	2e-38	unknown, similar to *Arabidopsis* hypothetical protein	BE195396	5e-23	115
711N16.15	<4045	>4466	+	CAA88558	7e-5	Glycine-rich RNA binding protein, pathogen response gene	BF256950	7e-4	27
711N16.14	9078	12450	-	P08640	8e-4	glucoamylase s1/s2 precursor	BE216561	7e-5	25
BPM	15085	21480	-	AAC95215 AAC95216	7e-50	putative pumilio/Mpt5 family protein [*Arabidopsis thaliana*]	AW982507 BG414539 AL511643	0 0 0	735 860 519
BMTa	<22721	>22906	-	AAF66185	2e-15	maturase [*Bromus inermis*]			
BMTb	<25134	>25530	-	AAF66186	4e-10	maturase [*Melica altissima*]			
BDI	<29965	>30140		AJ25065 (DNA only)	4e-50	induced in barleyby chemicals that activate disease resistance	BF260884 BE420733	1e-42 1e-41	176 158
711N16.9	37161	39002	+			unknown			
711N16.8	<50072	>52747	-	AF114171_8	1e-71	hypothetical protein [*Sorghum bicolor*]	BF617740	2e-65	167

Table 1. (continued)

Predicted Gene Product	Position[a]		Strand	BLASTX Search: Best Homolog GenBank ID	BLASTX Search: E-value	BLASTX Search: Predicted function[b]	EST BLASTn search: GenBank ID	EST BLASTn search: E-value	EST BLASTn search: Longest Match (bp)
RGH3b	<62862	>66015	-	AF149112	1e-44	stripe rust resistance protein Yr10 [*Triticum aestivum*]			
80H14.15b	74225	74797	+			unknown			
80H14.14b	75570	76794	+	AJ297951	1e-06	p23 co-chaperone [*Arabidopsis thaliana*]	BE196535	2e-42	194
RGH2b	77670	81459	+	AF207842	1e-98	Pi-ta protein [*Oryza sativa*]			
RGH1f	81517	85438	+	AF114171	7e-89	disease resistance protein RPM1 homolog [*Sorghum bicolor*]	BF065605	4e-18	96
80H14.11b	86562	87614	+	AF062655	5e-11	plenty-of-prolines-101 (POP101) [*Mus musculus*]			
80H14.10b	89999	93163				unknown			
RGH3a	102577	105820	-	AF149112	1e-44	stripe rust resistance protein Yr10 [*Triticum aestivum*]			
80H14.15a	113940	114509	+			unknown			
80H14.14a	115283	116507	+	AJ297951	1e-06	p23 co-chaperone [*Arabidopsis thaliana*]	BE196535	2e-42	194

Table 1. (continued)

Predicted Gene Product	Position[a]		Strand	BLASTX Search: Best Homolog GenBank ID	BLASTX Search: E-value	BLASTX Search: Predicted function[b]	EST BLASTn search: GenBank ID	EST BLASTn search: E-value	EST BLASTn search: Longest Match (bp)
RGH2a	117383	121172	+	AF207842	1e-98	Pi-ta protein [*Oryza sativa*]			
RGH1c	121230	125151	+	AF114171	7e-89	disease resistance protein RPM1 homolog [*Sorghum bicolor*]	BF065605	4e-18	96
80H14.11a	126275	127327	+	AF062655	5e-11	plenty-of-prolines-101 (POP101) [*Mus musculus*]			
80H14.10a	129712	132876	+			unknown			
RGH1bcd	185032	190492	-	AF149112	5e-83	stripe rust resistance protein Yr10 [*Triticum aestivum*]	BF065605	2e-43	122
80H14.8	194824	197941	+			unknown	AL502048 BG344356	3e-92 2e-90	220 341
CI2f	204440	204694	+	CAA40350	2e-39	chymotrypsin inhibitor 2 [*Hordeum vulgare*]	BE601601 BE196644 BG366350	0	391 399 411
CI2e	206954	207173	+	CAB71340	6e-20	putative proteinase inhibitor [*Hordeum vulgare*]	BF255218 BF253985	8e-27	154
RGH1a	207921	212117	+	AF149112	2e-88	stripe rust resistance protein Yr10 [*Triticum aestivum*]	BF065605	6e-14	57

Table 1. (continued)

Predicted Gene Product	Position[a]		Strand	BLASTX Search: Best Homolog GenBank ID	BLASTX Search: E-value	BLASTX Search: Predicted function[b]	EST BLASTn search: GenBank ID	EST BLASTn search: E-value	EST BLASTn search: Longest Match (bp)
Cl2d	217540	217764	-	CAB71340	5e-24	putative proteinase inhibitor [*Hordeum vulgare*]	BG300496	0	392
Cl2c	220610	220834	-	CAB71340	2e-32	putative proteinase inhibitor [*Hordeum vulgare*]	BF630148 BF629382 BF253985	0	501 414 399
Cl2b	223775	224008	-	CAB71340	2e-23	putative proteinase inhibitor [*Hordeum vulgare*]	BG300496 BF257305	1e-83	205
Cl2a	224982	225149	-	CAB71340	2e-10	putative proteinase inhibitor [*Hordeum vulgare*]	BF258541 BG300496	4e-43	109

[a]The carrots < > show the approximate sequence range.

[b]If the functional match is different from the best homolog, its GenBank accession number is given here.

Table 2. Base composition of three sequenced regions from the barley genome

Target Locus	Length (nt)	G+C (%)	CpG			Gene Island[a]		Gene Density (kb/gene)		Major retro.		Total retro (%)
			Ratio	Obs/Exp[b]	CpG/GpC	Length (kb)	%	Of the total sequence	Within gene islands	name	%	
Mla	261265	44.9	4.2%	0.83	1.21	~30/31[b]/41	51.0	8.2	4.4/4.4/4.6	*BARE-1*	17.5	40.2
Rar1	65979	46.0	4.2%	0.80	1.17	21	31.8	22.0	7	*BARE-1*	16.0	63.6
mlo	59748	45.6	4.2%	0.79	1.26	37	62.0	20.0	12	*BAGY-1*	24.1	41.7
Total	386992	45.2	4.2%	0.82	1.21	35	49.4	10.0		*BARE-1*	14.5	44.4

[a]The size of an gene island was calculated between two retro complexes or retroelements.

[b]The two duplication units was considered one gene island, but was used for percentage calculation.

[c]Observed vs. expected ratio of CpG content.

The *Mla* region contains multiple genes associated with plant defense

Multiple defense-related genes are present in the barley *Mla* region (Fig. 1*a* and Table 1). Eight Resistance Gene Homologs (*RGH*s) containing Coiled Coil (CC), NBS, and Leucine Rich Repeat (LRR) domains comprise the major defense-related group (Halterman et al., 2001). Members of the CC-NBS-LRR class are the most prevalent among cloned plant defense genes and are known to function in recognition of bacterial, fungal, viral, and nematode pathogens (Wise, 2000). The eight *RGH*s are grouped into three distinct families with less than 47% amino acid sequence similarity between the families, and 78-100% within families (Table 3). The only marked similarity between the three *RGH* families is the P-loop (GKTTL), kinase 2a (RYLVIIDDI), and the GGVPLA conserved domains found in all NBS-LRR encoding genes, which suggests functional similarity, but evolutionary independence.

RGH1a and *1e* were recently used to identify copies of the *Mla1* and *Mla6* resistance alleles that confer specificity to *Blumeria graminis* f. sp. *hordei* (*Bgh*), causal agent of the powdery mildew disease (Halterman et al., 2001; Zhou et al., 2001). Our results now indicate that *RGH1bcd* is the susceptible allele of *Mla1* and *Mla6* from cv. Morex. Besides sharing the highest sequence similarity with both the *Mla1* and *Mla6* resistance alleles, the *HORPIA-2* and *Alexandra* retrotransposons adjacent to *RGH1bcd* also flank the *Mla1* and *Mla6* sequences in the same orientation. During evolution of the *Mla* complex, intron 1 of *RGH1bcd* has been host to a *BARE-1* retrotransposon that has been excised through unequal recombination. Thus, the present day *RGH1bcd* contains only a solo LTR of *BARE-1* in intron 1 in addition to a 29-bp deletion in the LRR domain, resulting in premature termination of the open reading frame. *RGH1e* and *1f* reside in a 40-kb tandem duplication, and thus, both have the same mutation in amino acid 151 resulting in early stop codons in their NBS domains. Comparison of Morex *RGH1*-family homologs to known functional *Mla* alleles illustrates that almost all deletions have occurred in multiples of three nucleotides

Table 3. DNA and putative amino acid sequence similarities of the eight *Mla* resistance gene homologs (*RGH*) in the Morex *Mla* cluster[a]

	RGH designation						
RGH	*Mla6*	*RGH1a*	*RGH1bcd*	*RGH1e*	*RGH1f*	*RGH2a&b*	*RGH3a&b*
Mla6		*82*	*87*	*84*	*84*	*45*	*43*
RGH1a	78		84	83	83	46	44
RGH1bcd	84	81		86	86	47	44
RGH1e	79	79	81		100[b]	47	44
RGH1f	79	79	81	100		47	44
RGH2a&b	43	42	42	42	42		41
RGH3a&b	41	41	38	39	39	36	

[a] Comparisons were done with the GCG-GAP program. Numbers above diagonal markers designate nucleic acid similarities of coding region, whereas comparisons below the diagonal markers indicate predicted amino-acid similarities.

[b] There is a single nucleic acid sequence difference between *RGH1e* and *1f*, which results in no amino acid change.

(Fig. 2). This loss of nucleotide triplets would maintain a full-length open reading frame during evolution. Additionally, the lost amino acids are not conserved in the *Mla1* and *Mla6* functional alleles, suggesting that those deleted nucleotides are of little consequence for the overall function of MLA proteins and thus, the evolution of the *RGH1* family has been driven by selection for function. The *RGH3a* and *3b* pseudogenes contain an early stop codon in amino acid 582 of their LRR region, while *RGH2a* and *2b* are predicted to encode full-length CC-NBS-LRR proteins.

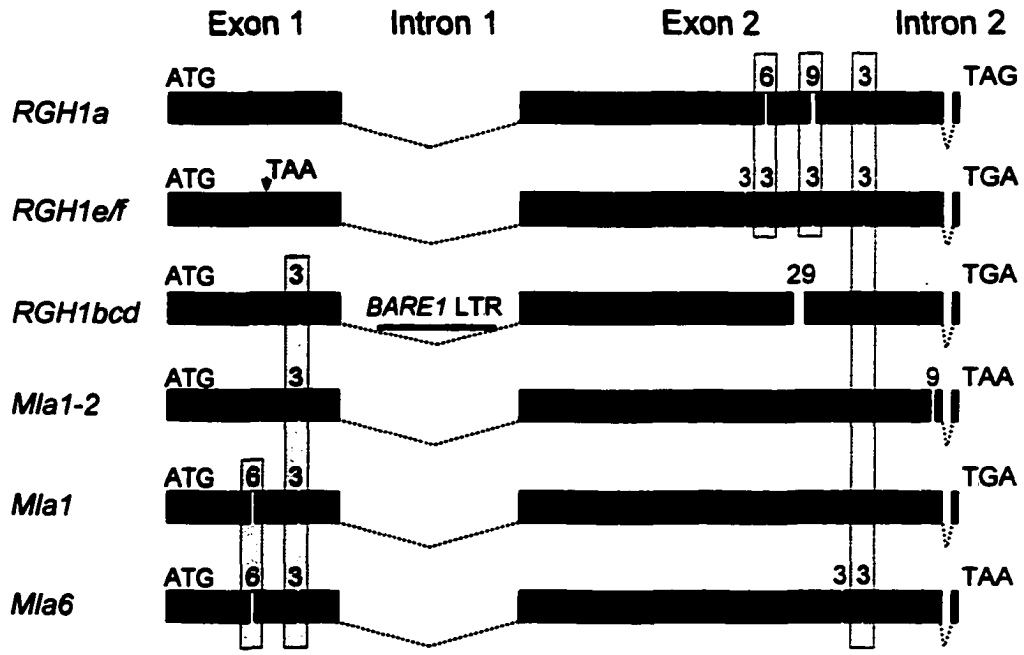


Fig. 2. Structural comparison of barley Morex *RGH1* family members with functional *Mla* specificities from accessions C.I. 16137 (*Mla1*) and C.I. 16151 (*Mla6*). Deletions in the coding regions are designated by white verticle bars and an insertion in intron 1 of *RGH1bcd* is designated by a solid line. The coding regions for *RGH1a*, *RGH1bcd* and *RGH1e&f* range from 2853 to 2892 bp. Except for *RGH1bcd*, the 1st intron ranges in size from 1068- to 1295-bp. The insertion of an 1836-bp *BARE-1* LTR into the 1st intron of *RGH1bcd* increases the size of this intron from 792- to 2628-bp. Intron 2 ranges in size from 108 to 123 bp. *RGH1e* and *1f* are both within the large fragment duplication and differ by only one nucleotide, resulting in no amino acid change. Above each white bar is the number of nucleotides for each deletion. Cv. Morex presently does not contain a characterized resistance specificity, however, since *RGH1a* appears to encode a full length CC-NBS-LRR gene, the nature of the gene-for-gene interaction does not preclude *RGH1a* from being functional, if the appropriate *Bgh* isolate were identified.

A family of six intact chymotrypsin inhibitor 2 (*CI2*) genes, range in size from 55 to 85 amino acids and flank *RGH1a* (Fig. 1 and Table 4). *CI2* genes belong to the potato inhibitor I family of serine protease inhibitors and function in various biochemical pathways, notably, to provide resistance to insect pathogens (Heath et al. 1997), wounding (Lee et al. 1986), and ethylene-regulated fruit ripening (Margossian et al. 1988). In addition, *CI2* genes can be induced by salicylic acid (SA) mimics and also by jasmonic acid (JA), resulting in elevated disease resistance in Gramineaeceous plants such as barley or maize (Cordero et al. 1994; Besser et al., 2000). In the *Mla* region, *CI2c* has been independently cloned as *BCI7* from cv. Ingrid, a 2,6-dichloroisonicotinic acid (DCINA; a SA mimic) induced gene, which provides enhanced resistance to *Bgh*. The predicted *BCI9*-like gene, *Bdi* (barley DCINA-induced) can also be induced by DCINA or JA, and concomitantly, barley plants show elevated resistance to *Bgh* (Besser et al., 2000). Finally, the glycine-rich 711N16.15 protein (top of Fig. 1) is similar to HvGRP3, a glycine-rich RNA-binding PR protein (Molina et al., 1997). Neither *BCI7* nor *BCI9* can be induced only by inoculation with *Bgh*. However, *HvGRP3* is inducible by *Bgh* in both compatible and incompatible interactions (Molina et al., 1997).

The last complete gene in the *Mla* cluster is *Bpm*, a barley *Pum/Mpt5/FBF*-like gene. The products of *Pum/FBF/Mpt5* genes (*Puf* family) have been shown to repress mRNA translation by binding to 3' UTR of RNAs in *Drosophila*, *C. elegans*, and yeast (Tadauchi, et al. 2001). In plants, putative *Puf* orthologs have been reported only in *Arabidopsis* (Lin et al., 2000), but their function in plants is yet untested.

The *Mla* region contains all major classes of transposable elements

All major classes of transposable elements are present in the *Mla* region (Fig. 1a, b and c). Among them, the largest class in the region is the LTR (long terminal repeat) class of retrotransposons, including four *copia*-like families (*BARE-1*, *HORPIA-1* to *3*), three *gypsy*-

Table 4. DNA and putative amino-acid similarities of six chymotrypsin inhibitor genes (*CI2*) flanking *RGH1a* [a].

	CI2 designation					
CI2	*CI2a*	*CI2b*	*CI2c*	*CI2d*	*CI2e*	*CI2f*
CI2a		87	74	75	65	57
CI2b	85		80	88	69	57
CI2c	69	77		84	77	57
CI2d	71	85	78		72	56
CI2e	62	68	71	61		59
CI2f	62	61	70	61	58	

[a] Comparisons were done with the GCG-GAP program. Numbers above diagonal markers designate nucleic acid similarities of coding region, whereas comparisons below the diagonal markers indicate predicted amino-acid similarities.

like families (*BAGY-2*, *HORGY-1*, and *Sukkula*) and one *Athila*-like element (*Sabrina*). *Sukkula* and *Sabrina* were previously reported only as solo LTRs (Shirasu et al., 2000). This report represents the first description of the full-length elements and thus enables their classification into the *Gypsy* and *Athila* families, respectively. *BARE-1* (Manninen and Schulman, 1993) is the major retrotransposon, accounting for 17.5% of the sequence. Five of the seven *BARE-1* elements are full length (Fig. 1b, 1c). Of these five, point mutations resulting in premature stop codons interupt the polyprotein open reading frames. Therefore, the transposition of these compromised *BARE-1* elements is most likely dependent upon a

limited number of fully functional *BARE-1* or other related elements. The remaining two (solo LTR) *BARE-1* elements are a consequence of intra-element recombination as confirmed by their 5-bp (direct repeat) insertion signatures.

Eight MITE (Miniature Inverted Transposable Element) families that range in size from 29 to 528 bp (Table 5) are positioned near the 5' or 3' end of predicted genes. As the most abundant transposable MITE family, eight *Stowaway* elements (Bureau and Wessler, 1994) are present in the region, ranging from 128 to 243 bp in size. The 16-member, 39-bp *HORMITE-3* family possesses a unique clustered arrangement of three to five members interspersed with short sequences. This tandem arrangement predicts that the transposable unit could be flexible in size.

An apparent full-length, 6,459-bp *MuDR*-like transposon, *HORMU-1*, is inserted into the *Stowaway.4* MITE element. Its 179-bp terminal inverted repeats are 94% sequence similar and its pseudogene encodes a 654-aa transposase with an early stop codon at aa position 489. Southern analysis using an internal portion of the transposase-encoding sequence demonstrated that *HORMU-1* has at least 50 copies in the barley genome (data not shown). *Mutator* transposons are widespread in grasses (Lisch et al. 2001), thus, it is likely that there is an intact *HORMU-1* element in the barley genome that still possesses transposase activity. A 3,578-bp, new non-LTR LINE element, designated *Alexandra*, is inserted into *Stowaway.7*, and was identified via Blastx search for polyprotein, its 3' polyA signal and 7-bp ATACGAA insertion signature. Finally, a truncated Ac-like transposon was also located in the region next to the *BCI-9*-like gene. Taken together, transposon-like sequences comprise 44.9% of the sequence with retroelements accounting for 40.2%, transposons for 3.6%, and MITE elements for 1.1%. Fifty-four simple sequence repeats (SSR) were identified in the region (Table 6). Sixteen of them are associated with retro elements, four are associated with transposons, and five are associated with genes. The remaining twenty-

Table 5. MITE elements in the barley *Mla* region

MITE Element	Position		Size (bp)	MITE Element	Position		Size (bp)
Hormite1				*Stowaway*			
Hormite1	29590	29746	157	*Stowaway.1*	148	309	162
Hormite2				*Stowaway.2*	1592	1754	163
Hormite2	31684	32211	528	*Stowaway.3*	40068	40310	243
Hormite3				*Stowaway.4*	142691	142853	163
Hormite3.1	49087	49120	34	*Stowaway.5*	136684	136726	158
Hormite3.2	49126	49163	38		143230	143342	
Hormite3.3	49203	49241	39	*Stowaway.6*	217169	217348	181
Hormite3.4	54772	54810	39	*Stowaway.7*	190752	190812	164
Hormite3.5	54815	54853	39		194398	194500	
Hormite3.6	54856	54894	39	*Stowaway.8*	222790	222917	128
Hormite3.7	54932	54970	39	*Hormite4*			
Hormite3.8	54974	55012	39	*Hormite4*	55142	55249	108
Hormite3.9	61214	61252	39	*Hormite5*			
Hormite3.10	61297	61335	39	*Hormite5.1*	73203	73231	29
Hormite3.11	61380	61418	39	*Hormite5.2*	112918	112946	29
Hormite3.12	61460	61498	39	*Hormite6*			
Hormite3.13	100929	100967	39	*Hormite6*	203605	203667	63
Hormite3.14	101012	101050	39	*Hormite7*			
Hormite3.15	101095	101133	39	*Hormite7*	192191	192377	187
Hormite3.16	101175	101213	39				

Table 6. Simple Sequence Repeats in the barley *Mla* region

Sequence Name	Position		Length (bp)	Strand[a]	Repeat Name	Repeat Type[b]
	Start	End				
Morex_Mla	4290	4365	76	+	(GGGGA)n	Simple_repeat
Morex_Mla	4572	4592	21	C	(GGAA)n	Simple_repeat
Morex_Mla	4715	4734	20	+	(TAG)n	Simple_repeat
Morex_Mla	10155	10246	92	C	(GAAAA)n	Simple_repeat
Morex_Mla	12816	12904	89	C	(TAAAA)n	Simple_repeat
Morex_Mla	13195	13223	29	C	AT_rich	Low_complexity
Morex_Mla	21422	21505	84	C	(CGG)n	Simple_repeat
Morex_Mla	21537	21567	31	+	(CAAAC)n	Simple_repeat
Morex_Mla	29149	29170	22	C	GC_rich	Low_complexity
Morex_Mla	35111	35203	93	+	(CGG)n	Simple_repeat
Morex_Mla	35838	35858	21	+	(CGA)n	Simple_repeat
Morex_Mla	41662	41755	94	C	(GAAAA)n	Simple_repeat
Morex_Mla	41765	41793	29	C	(TAAAAA)n	Simple_repeat
Morex_Mla	41839	42008	170	C	(TAAAAA)n	Simple_repeat
Morex_Mla	41935	42041	107	C	(GAAA)n	Simple_repeat
Morex_Mla	42601	42637	37	C	AT_rich	Low_complexity
Morex_Mla	43972	44040	69	C	(TAAA)n	Simple_repeat
Morex_Mla	45140	45309	170	C	(CGGGGG)n	Simple_repeat
Morex_Mla	47534	47560	27	C	(GA)n	Simple_repeat
Morex_Mla	50194	50240	47	C	(CA)n	Simple_repeat
Morex_Mla	53934	54111	178	C	(GAA)n	Simple_repeat
Morex_Mla	54382	54460	79	C	(GAA)n	Simple_repeat
Morex_Mla	54505	54650	146	+	(CGG)n	Simple_repeat
Morex_Mla	61081	61147	77	+	(GAGAA)n	Simple_repeat
Morex_Mla	66106	66126	21	C	(CGAGG)n	Simple_repeat
Morex_Mla	66546	66619	74	C	AT_rich	Low_complexity
Morex_Mla	66661	66689	29	C	AT_rich	Low_complexity

Table 6. (continued)

Sequence Name	Position		Length (bp)	Strand[a]	Repeat Name	Repeat Type[b]
	Start	End				
Morex_Mla	66693	66759	67	C	(CGA)n	Simple_repeat
Morex_Mla	69826	69870	45	+	(GGA)n	Simple_repeat
Morex_Mla	74065	74086	22	+	(CA)n	Simple_repeat
Morex_Mla	75492	75516	25	C	(GA)n	Simple_repeat
Morex_Mla	76707	76782	76	+	(GGA)n	Simple_repeat
Morex_Mla	86793	86833	41	C	(AGGGGG)n	Simple_repeat
Morex_Mla	86894	87075	182	+	(CCCCG)n	Simple_repeat
Morex_Mla	94902	94924	23	C	GC_rich	Low_complexity
Morex_Mla	97560	97639	80	C	(GA)n	Simple_repeat
Morex_Mla	98305	98366	62	C	(TA)n	Simple_repeat
Morex_Mla	100796	100862	67	+	(GAGAA)n	Simple_repeat
Morex_Mla	105821	105841	21	C	(CGAGG)n	Simple_repeat
Morex_Mla	106261	106334	74	C	AT_rich	Low_complexity
Morex_Mla	106376	106404	29	C	AT_rich	Low_complexity
Morex_Mla	106408	106474	67	C	(CGA)n	Simple_repeat
Morex_Mla	109541	109585	45	+	(GGA)n	Simple_repeat
Morex_Mla	113780	113801	22	+	(CA)n	Simple_repeat
Morex_Mla	115205	115229	25	C	(GA)n	Simple_repeat
Morex_Mla	116420	116495	76	+	(GGA)n	Simple_repeat
Morex_Mla	126506	126546	41	C	(AGGGGG)n	Simple_repeat
Morex_Mla	126607	126788	182	+	(CCCCG)n	Simple_repeat
Morex_Mla	134615	134637	23	C	GC_rich	Low_complexity
Morex_Mla	137274	137353	80	C	(GA)n	Simple_repeat
Morex_Mla	138019	138074	56	C	(TA)n	Simple_repeat
Morex_Mla	154282	154401	120	+	(CGG)n	Simple_repeat
Morex_Mla	155559	155585	27	+	(G)n	Simple_repeat
Morex_Mla	156478	156596	119	+	(CGGGGG)n	Simple_repeat

Table 6. (continued)

Sequence Name	Position		Length (bp)	Strand[a]	Repeat Name	Repeat Type[b]
	Start	End				
Morex_Mla	156662	156846	185	+	(CGGGGG)n	Simple_repeat
Morex_Mla	156748	156925	178	C	(CACCC)n	Simple_repeat
Morex_Mla	156893	157076	184	+	(CGGGGG)n	Simple_repeat
Morex_Mla	157499	157524	26	C	AT_rich	Low_complexity
Morex_Mla	157955	158014	60	+	(CAGAG)n	Simple_repeat
Morex_Mla	181488	181584	97	C	(CGA)n	Simple_repeat
Morex_Mla	188509	188572	64	+	(GAGAA)n	Simple_repeat
Morex_Mla	192124	192155	32	C	AT_rich	Low_complexity
Morex_Mla	193878	193898	21	C	(G)n	Simple_repeat
Morex_Mla	195753	195782	30	C	AT_rich	Low_complexity
Morex_Mla	199811	199834	24	C	AT_rich	Low_complexity
Morex_Mla	200835	200872	38	C	AT_rich	Low_complexity
Morex_Mla	201097	201125	29	C	AT_rich	Low_complexity
Morex_Mla	202516	202567	52	C	AT_rich	Low_complexity
Morex_Mla	203458	203514	57	C	AT_rich	Low_complexity
Morex_Mla	203559	203586	28	C	AT_rich	Low_complexity
Morex_Mla	203712	203748	37	C	AT_rich	Low_complexity
Morex_Mla	219736	219780	45	C	AT_rich	Low_complexity
Morex_Mla	221418	221448	31	C	AT_rich	Low_complexity
Morex_Mla	222677	222729	53	C	AT_rich	Low_complexity
Morex_Mla	230915	230972	58	C	(GAGAA)n	Simple_repeat
Morex_Mla	230933	230973	41	C	(GAAAA)n	Simple_repeat
Morex_Mla	242584	242611	28	C	AT_rich	Low_complexity

[a] + plus strand, C minus strand.

[b] Among the 77 events, 20 are AT_rich, 3 are GC_rich, and 54 are SSR. Twenty of the SSR are within the retroelement regions, five in genes, and five in duplicated region.

nine are in intergenic regions and occupy 1% of the genomic sequence. We did not observe a higher density of SSRs in retrotransposons as previously reported (Ramsay et al., 1999).

Molecular evolution of the *Mla* region provides evidence for genome expansion in barley

Due to the mechanism of transposition, the two LTRs of a retrotransposon can be assumed to be identical at the time of their insertion. Thus, calculation of the base substitution rate between the two LTRs reflects the relative age of the insertion event. We can utilize the average base substitution rate derived from the grasses (Poaceae) *adh1-adh2* region of 6.5 x 10^{-9} per site per year (Gaut et al. 1996; SanMiguel et al. 1998) as a general guideline with the assumptions of uniform mutation rates and no gene conversion. Using this figure combined with the calculated base substitution rate in Table 7, the time of insertion of the five *BARE-1* elements is between 0.6 to 2 million years. Except for the *HORGY-1* element inserted to the region about 7.3 million years ago, all other retroelements were inserted less than 2.5 million years, which is comparable to less than three million years of most retrotransposon insertions in maize (SanMiguel et al. 1998).

Figure 3 illustrates a model for the evolution of the *Mla* region. In this model, a 30-kb progenitor sequence, including *RGH1bcd*, *CI2d*, and *CI2c*, sustained duplication and inversion events to create *CI2f*, *C2Ie*, and *RGH1a*. Following these events, a second duplication and inversion of *RGH1bcd* gave rise to *RGH1e*, plus *CI2a* arose from a duplication of *CI2b*. Following the gene duplications and inversions described above, our data indicate that the "founder" element *HORGY-1* of Complex I inserted between *RGH1bcd* and *RGH1e*, and subsequently acted as a receiver for a total of 44,595-bp of nested transposons (Complex I, Fig. 1b). Seven additional insertion events took place, including MITE *Stowaway.5* into retro *HORGY-1*, transposon *HORMU-1* into *Stowaway.5*, and

Table 7. Estimated time of retrotransposon insertion in the *Mla* region

Element[a]	Length of LTR[b]	Total substitution[b]	Base substitution rate[c]	Time (myrs)[d]
BARE-1b	1813	15	0.0083	0.64
BARE-1c	1717	20	0.0116	0.89
Sabrina	1526	48	0.0315	2.42
BARE-1f	1817	22	0.0121	0.93
BARE-1d	1801	42	0.0233	1.79
BARE-1e	1761	47	0.0267	2.05
HORGY-1	286	27	0.0944	7.26
HORPIA-2	321	2	0.0062	0.48
HORPIA-3	521	14	0.0269	2.07

[a] Full length retrotransposon with two LTRs

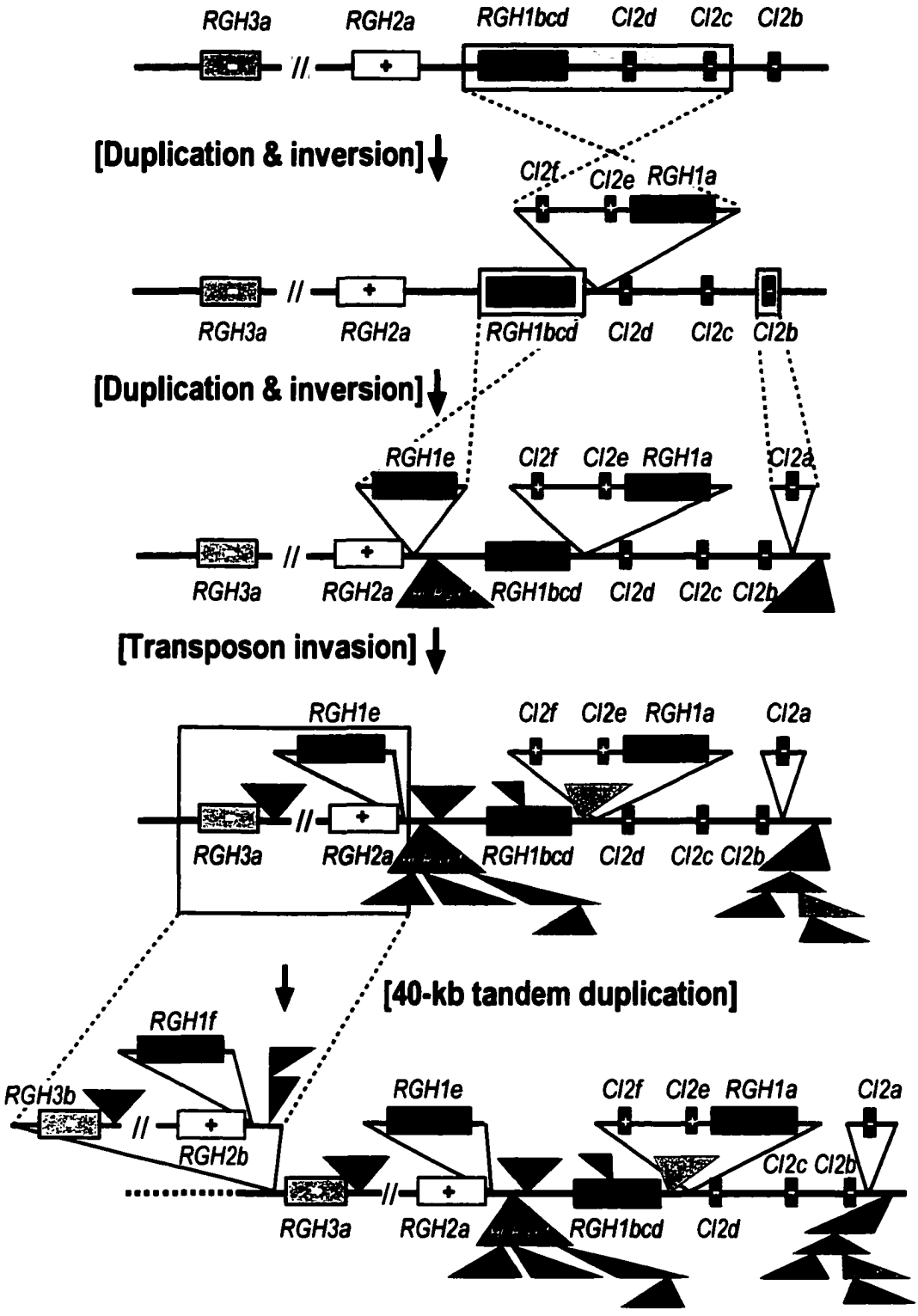
[b] Deletion or insertion was counted as a single event including a retrotransposon insertion. The substitution was calculated as the total nucleotide difference between the two LTRs.

[c] For less than 10% base substitution rate, the rate can be used as an estimated number of substitutions per nucleotide (k) for calculation of divergence time

[d] Divergence time as calculated as $k/2*k_{nus}$. k_{nus}designates the base substitution rate per nucleotide per year.

Stowaway.4 into *HORMU-1*. Similarly, in Complex II (Fig. 1c), *HORPIA-1* functioned as a receiver for five retrotransposon insertion events totaling 35,945 bp. The total nucleic acid sequence of the two nested complexes is 87,586 bp, which comprise 33.5% of the total sequence. All of the retro elements in complex I and II have both LTRs and polyproteins. This is in contrast to the nested insertion of only solo LTRs in the barley *Rar1* region (Shirasu et al., 2000), but consistent with retroelement structure at the maize *Adh1* locus

Fig. 3. Evolution of the barley *Mla* region. Of the 34 predicted genes, 24 have at least one homolog. The *RGH1* and *CI2* families represent the major gene duplications in the region and account for 16% of the sequence. The remaining gene duplications have arisen from tandem fragment duplications, of which 14 are located in the 40-kb, gene-rich tandem repeat and two in the 2.6-kb tandem repeat. The DNA sequence similarities of *RGH1bcd*, *CI2d* and *CI2c* as compared to *RGH1a*, *CI2f* and *CI2e* in the opposite orientation indicate an ancient inversion event followed by divergent evolution. A more recent duplication and inversion to create *RGH1e* is indicated by the higher sequence similarity between *RGH1bcd* and *RGH1e* than that between *RGH1bcd* and *RGH1a* (Table 3). Similarly, *CI2a* possesses the highest sequence similarity with *CI2b* within the *CI2* family (Table 4). Genes are shown in rectangles with different colors representing different families, while transposable elements designated by triangles. The colors match with Fig. 1 for each element. + or – indicates the transcriptional direction of the related genes, +: forward strand, -: complementary strand.


RGH3a
RGH2a
RGH1bcd
Cl2d
Cl2c
Cl2b
[Duplication & inversion]
Cl2f
Cl2e
RGH1a
[Duplication & inversion]
RGH1e
Cl2a
[Transposon invasion]
[40-kb tandem duplication]
RGH1f
RGH3b
RGH2b

(SanMiguel et al., 1996, Tikhonov et al., 1999). However, in the *adh1* locus, the nested elements are only retrotransposons, while there are transposon and MITE elements in addition to retrotransposons at the *Mla* locus. Following this *HORGY-1*-based transposon invasion, a final duplication of 40-kb of gene-rich sequence expanded the original 30-kb, including the ancient *Mla* allele, *RGH1bcd*, to the over 200-kb sequence we see today.

Discussion

Functional clustering of defense-related genes was found in the barley *Mla* complex. Of the 32 protein-coding genes detected in the *Mla* region, there are sixteen defense-related genes that can be classified into six families. These include three CC-NBS-LRR *RGH* families, one *CI2* family, one chemically-induced family and one PR-protein family. Four of the six families function in three independent pathways related to powdery mildew resistance, including the *Mla* gene-for-gene pathway, an induced systemic resistance (ISR) pathway, and the PR protein pathway. The *Mla* signal cascade has been reported to be SA-independent in the incompatible *Mla-Bgh* interaction (Huckelhoven et al., 1999). In contrast, *BCI-7* of the six-member *CI2* family and the *BCI-9*-like gene function via SA induction in the ISR pathway. Both of these genes are also induced by JA, resulting in systemic resistance to *Bgh*. More interestingly, *BCI-7* and *BCI-9* can not be induced by direct interaction of barley with *Bgh*, indicating that chemically induced systemic resistance to *Bgh* operates via a different pathway than gene-for-gene specificity (Besser et al., 2000). Finally, there is an unknown PR-protein related pathway, in which the glycine-rich *711N16.15* can be induced by both compatible and incompatible *Bgh*-barley interactions.

Our finding of defense-gene clustering appears to be quite prevalent in plant systems. For example, *Tip1* and *Tip2* are two additional genes adjacent to barley *Rar1*, a gene required

for some, but not all, *Mla* specified resistance (Shirasu, et al. 2000). *Tip* homologs in *Arabidopsis* are inducible by both SA and JA (Schenk et al., 2000). We might predict that *Tip* and *Rar1* genes would function in two different pathways, similar to *CI2* and *Mla* genes. Adjacent to the barley *Mlo* (broad spectrum powdery mildew resistance) gene is a ring finger protein (Panstruga et al. 1998); the tobacco homolog of this gene can be rapidly elicited in *Avr9/Cf9* interactions (GenBank Accession AAG43550). In this regard, it would be interesting to learn the functional relationship of the membrane-anchored MLO and membrane-bound AVR9/CF9 interaction (Piedras et al., 2000). On *Arabidopsis* chromosome 4 (GB accession AL161551), a 180-kb region encodes six *RGHs* of the NBS-LRR class with less than 35% aa identity, a nine-membered PR-2 (chitinase) gene family, a *NPR1* homolog, and several potential SA or JA inducible gene families. The *NPR1* gene is a major signal-transduction component of NBS-LRR resistance pathway and required for SA-dependent *R* gene function (Cao et al., 1997). Furthermore, there is a 40-kb region on *Arabidopsis* chromosome 1 (GB accession U95973) that contains an *RPM1* disease resistance gene homolog, an *mlo* resistance gene homolog, and two distinct serine/threonine kinase genes. There is also a serine/threonine kinase pseudogene next to *RPP5* in Arabidopsis (Noel et al., 1999) and six tandem serine/throenine kinase *Pto* paralogs adjacent to the *Prf* gene, required for *Pto*-specified resistance to bacterial speck disease in tomato (Jia et al., 1997).

The 8-Mb, mammalian major histocompatibility complex (MHC) also exhibits functional clustering of defense-related genes (Beck and Trowsdale, 2000). About 40% of the genes in the MHC class I are immunity-related. All genes in the class II region have immune functions, including class II A and B genes, low molecular weight proteins (LMPs), transporters associated with antigen processing (TAPs), and TAP binding proteins (TAPBP). In the class III region, over seven genes are involved in inflammation response, including the

tumour necrosis factor (TNF) super family. Functional clustering is similar to that of prokaryotic organisms and may indicate that defense is one of the oldest life functions.

Materials and Methods

DNA template preparation and sequencing

Shotgun sequencing libraries were constructed as described previously by ligating randomly sheered BACs into the pUC18 vector (Wei et al., 1999). DNA sequence templates were prepared by the alkali lysis method (Sambrook et al., 1989) with minor modification. *E. coli* cells were incubated in 96-well culture blocks with 1.2 ml/well TB medium [1.2% bacto-tryptone, 2.4% bacto-yeast extract, 0.4% glycerol (v/v), 0.017 M KH_2PO_4, 0.072 M K_2HPO_4] at 37 °C for at least 16 h at 325 rpm. Subsequent to centrifugation, cell pellets were re-suspended in 100 µl solution I [50 mM glucose, 10 mM EDTA (pH 8.0) and 25 mM Tris-HCl (pH 8.0), 0.2 mg RNase A], lysed with 100 µl solution II (0.2 N NaOH, 1% SDS), and neutralized in 100 µl solution III (3 M KOAc, pH 4.8). The lysate was transferred to 96-well MultiScreen-NA filter plates (Millipore, Bedford, MA) positioned on 96-well receiver plates (Whatman Polyfiltronics, Rockland, MA). The paired plates were centrifuged at 4000 rpm (4 °C) for 15 min to filter cell debris, and DNAs were precipitated by adding 200 µl isopropanol at RT for 10 min and centrifuged at 4000 rpm (4 °C) for 15 min. All of the above solutions were mixed by gently vortexing. Sequence-ready DNA pellets were washed with 70% ethanol, air-dryed for 10 min, and dissolved in 40 µl sterile ddH_2O.

Sequence data were obtained with universal M13 primers by using ABI BigDye terminators (PE Biosystems) at the Clemson University Genomics Institute DNA Sequencing Center (CUGI) and DNA Sequencing and Synthesis Facility (DSSF) at Iowa State University.

Sequence assembling, quality assessment and contig gap filling

Sequence data were assembled with the Phred and Phrap software package (Ewing et al., 1998; Ewing and Green, 1998) and edited in the Consed program (Gordon et al., 1998). At seven to eight times redundancy, internal low quality regions of each contig were identified and treated as gaps or different contigs. To fill gaps in the sequence, end clones of contigs were sequenced with the M13 reverse primer. These data were used for re-assembling, and the above protocol was repeated two more times. For further gap filling, a primer walking procedure was used for sequencing clones, whose sequences are located at the end of different contigs. Sequencing primers were designed by MacVector software package (Oxford Molecular Group, Inc., Campbell, CA) and synthesized at the DSSF of Iowa State University. GTG primer Mix (PE Biosystem) was added to resolve the potential secondary structure in remaining gaps consisting AT- or GC-rich regions. Virtual restriction enzyme digests of the final sequences were compared with *Hin*dIII, *Eco*RI, and *Mlu*I digests for a final confirmation that the sequence assemblies were correct.

Sequence annotation and computational analysis

Web-based Programs GENSCAN (Burge and Karlin, 1997), and GeneMarker (Lukashin and Borodovsky, 1998) were used for general gene mining. BLASTx, n and p of NCBI (National Center for Biotechnology Information, Altschul et al., 1997) were used to identify known genes or repetitive elements at intervals of 500 bp for BLASTx and 1000 bp for BLASTn. The full length of these BLAST-identified genes were determined by ORF finder (http://www.ncbi.nlm.nih.gov /gorf/gorf.html) and the BLAST aligned results. The LTR class of retrotransposons was identified by their LTRs, the inverted repeats of the LTRs and their insertion signatures. A DotPlot program was used for self-comparison to determine MITE elements at a 600-bp interval with a 400-bp rolling window (http://alces.med.umn.edu/rawdot.html, University of Minnesota). SSRs were analyzed by

RepeatMasker program (A.F.A. Smit & P. Green, unpublished data: http://repeatmasker.genome.washington.edu/cgi-bin/RM2_req.pl). The Wisconsin GCG package (Oxford Molecular, Madison, WI) GAP program was used for similarity analysis, and MacVector was used for translation analysis.

Acknowledgements

The authors thank D. Voytas and S. Whitham for critical review of the manuscript, X. Gu for helpful advice for the paleontology analysis, and L. Mao, S-S. Woo, M. Sasinowski, Y. Yu, R. Kingsbury, T. Rambo and the staff in the CUGI Bioinformatics and Sequencing Centers for technical assistance. This research was supported in part by USDA-NRI/CGP grants 98-35300-6169 & 00-35300-9213 to RPW and the NSF MRI #9724557 to RAW. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies. Joint contribution of the Corn Insects & Crop Genetics Research Unit, USDA-Agricultural Research Service and the Iowa Agriculture and Home Economics Experiment Station. This is Journal Paper No. J-19346 of the Iowa Agricultural and Home Economics Experiment Station, Ames, IA., Project No. 3368, and supported by Hatch Act and State of Iowa funds.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389-3402 (1997).

Baker, B., Zambriski, P., Staskawicz, B., and Dinesh-Kumar, S.P. Signaling in plant-microbe interactions. *Science* **276**, 726-733 (1997).

Beck, S., and Trowsdale, J. The human major histocompatibility complex: lessons from the DNA sequence. *Annu. Rev. Genom. Hum. Genet.* **1**, 117-137 (2000).

Besser, K., Jarosch, B., Langen, G., and Kogel, K.H. Expression analysis of genes induced in barley by chemical activation reveals distinct disease resistance pathways. *Mol. Plant Pathol.* **1**, 277-286 (2000).

Bureau, T.E., and Wessler, S.R. *Stowaway*: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *Plant Cell* **6**, 907-16 (1994).

Burge, C., and Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78-94 (1997).

Cao, H., Glazebrook, J., Clarke, J.D., Volko, S., and Dong, X. The Arabidopsis *NPR1* gene that controls systemic acquired resistance encodes a novel protein containing ankyrin repeats. *Cell* **88**, 57-63 (1997).

Cordero, M.J, Raventos, D., and San Segundo, B. Expression of a maize proteinase inhibitor gene is induced in response to wounding and fungal infection: systemic wound-response of a monocot gene. *Plant J.* **6**, 141-150 (1994).

Ewing, B., and Green, P. Basecalling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186-194 (1998).

Ewing B., Hillier, L., Wendl, M., and Green, P. Basecalling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175-185 (1998).

Gordon, D., Abajian, C., and Green, P. Consed, A graphical tool for sequence finishing. *Genome Res.* **8**, 195-202 (1998).

Gaut, B.S., Morton, B.R., McCaig, B.C., and Clegg, M.T. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel

rate differences at the plastid gene *rbcL*. *Proc. Natl. Acad. Sci. USA* **93**, 10274-10279 (1996).

Halterman, D., Zhou, F., Wei, F., Wise, R.P., and Schulze-Lefert, P. The MLA6 Coiled-Coil, NBS-LRR protein confers *AvrMla6*-dependent resistance specificity to *Blumeria graminis* f. sp. *hordei* in barley and wheat. *Plant J.* **25**, 335-348 (2001).

Hammond-Kosack, K.E., and Jones, J.D.G. Plant disease resistance genes. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48**, 575-607 (1997).

Heath, R.L., McDonald, G., Christeller, J.T., Lee, M., Bateman, K., West, J., Van Heeswijck, R., and Anderson, Marilyn, A. Proteinase inhibitors from *Nicotiana alata* enhance plant resistance to insect pests. *J. Insect Physiol.* **43**, 833-842 (1997).

Huckelhoven, R., Fodor, J., Preis, C., and Kogel, K.H. Hypersensitive cell death and papilla formation in barley attacked by the powdery mildew fungus are associated with hydrogen peroxide but not with salicylic acid accumulation. *Plant Physiol.* **119**, 1251-1260 (1999).

Jia, Y., Loh, Y.T, Zhou, J., Martin, G.B. Alleles of *Pto* and *Fen* occur in bacterial speck-susceptible and fenthion-insensitive tomato cultivars and encode active protein kinases. *Plant Cell 9*, 61-73 (1997).

Lee, J.S., Brown, W.E., Graham, J.S., Pearce, G., Fox, E.A., Dreher, T.W., Ahern,K.G., Pearson, G.D. and Ryan, C.A. Molecular characterization and phylogenetic studies of a wound-inducible proteinase inhibitor I gene in Lycopersicon species. *Proc. Natl. Acad. Sci. USA* **83**, 7277-7281 (1986).

Lin, X., Kaul, S., Rounsley, S., Shea, T.P., Benito, M.I., et al. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* **402**, 761-768 (1999).

Lisch, D.R., Freeling, M., Langham, R.J., and Choy, M.Y. *Mutator* transposase is widespread in the grasses. *Plant Physiol.* **125**, 1293-1303 (2001).

Lukashin, A.V., and Borodovsky, M. GeneMark.hmm: new solutions for gene finding. *Nucl. Acids Res.* **26**, 1107-1115 (1998).

Margossian, L.J., Federman, A.D., Giovannoni, J.J. and Fischer, R.L. Ethylene-regulated expression of a tomato fruit ripening gene encoding a proteinase inhibitor I with a glutamic residue at the reactive site. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8012-8016 (1988).

Manninen, I., and Schulman, A.H. BARE-1, a copia-like retroelement in barley (*Hordeum vulgare* L.). *Plant Mol. Biol.* **22**, 829-846 (1993).

Molina, A., Mena, M., Carbonero, P., and Garcia-Olmedo, F. Differential expression of pathogen-responsive genes encoding two types of glycine-rich proteins in barley. *Plant Mol. Biol.* **33**, 803-810 (1997).

Moore, G. Cereal chromosomes structure, evolution, and pairing. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **51**, 195-223 (2000).

Noel, L., Moores, T.L., van Der Biezen, E.A., Parniske, M., Daniels ,M.J., Parker, J.E., and Jones, J.D. Pronounced intraspecific haplotype divergence at the *RPP5* complex disease resistance locus of Arabidopsis. *Plant Cell* **11**, 2099-2112 (1999).

Panstruga, R., Buschges, R., Piffanelli, P. & Schulze-Lefert, P. A contiguous 60 kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. *Nucleic Acids Res.* 26, 1056-1062 (1998).

Piedras, P., Rivas, S., Droge, S., Hillmer, S, and Jones, J.D. Functional, c-myc-tagged Cf-9 resistance gene products are plasma-membrane localized and glycosylated. *Plant J.* **21**, 529-536 (2000).

Ramsay, L., Macaulay, M., Cardle, L., Morgante, M., degli Ivanissevich, S., Maestri, E., Powell, W., and Waugh, R. Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. *Plant J.* **17**, 415-425 (1999).

Sambrook, J., Fritsch, E.F., and Maniatis, T. *Molecular cloning : a laboratory manual.* 2nd ed. (1989) (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).

SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S.,Edwards, K.J., Lee, M., Avramova, Z., and Bennetzen, J.L. Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**. 765-768 (1996).

SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., and Bennetzen, J.L. The paleontology of intergene retrotransposons of maize. *Nat. Genet.* **20**, 43-45 (1998).

Schenk, P.M., Kazan, K., Wilson, I., Anderson, J.P., Richmond, T., Somerville, S.C., and Manners, J.M. Coordinated plant defense responses in Arabidopsis revealed by microarray analysis. *Proc. Natl. Acad. Sci. USA* **97**, 11655-11660 (2000).

Shirasu, K., Schulman, A.H., Lahaye, T., and Schulze-Lefert, P., A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. *Genome Res.* **10**. 908-915 (2000).

Tadauchi, T., Matsumoto, K., Herskowitz, I., and Irie, K. Post-transcriptional regulation through the HO 3'-UTR by *Mpt5*, a yeast homolog of *Pumilio* and *FBF*. *EMBO J.* **20**, 552-561 (2001).

Tikhonov, A.P., SanMiguel, P.J., Nakajima, Y., Gorenstein, N.M., Bennetzen, J.L., and Avramova, Z. Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. *Proc. Natl. Acad. Sci. USA* **96**, 7409-7414 (1999).

Van der Biezen, E.A., and Jones, J.D. The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. *Curr. Biol.* **8**, R226-R227 (1998).

Wei, F., Gobelman-Werner, K., Morroll, S.M., Kurth, J., Mao, L., Wing, R., Leister, D., Schulze-Lefert, P., and Wise, R.P. The *Mla* (powdery mildew) resistance cluster is

associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. *Genetics* **153**, 1929-1948 (1999).

Wise, R. P. Disease resistance: What's brewing in barley genomics. *Plant Disease* **84**, 1160-1170 (2000).

Zhou, F., Kurth, J., Wei, F., Elliott, C., Vale, G., Yahiaoui, N., Keller, B., Somerville, S., Wise, R.P. and Schulze-Lefert, P. Cell-autonomous expression of barley *Mla1* confers race-specific resistance to the powdery mildew fungus via a *Rar1*-independent signaling pathway. *Plant Cell* **13**, 337-350 (2001).

CHAPTER 4. GENES ARE TRANSCRIPTIONALLY ACTIVE WITHIN A HETEROCHROMATIC KNOB-LIKE SEQUENCE

A paper to be submitted to Genome Research

Fusheng Wei[1,2] and Roger P. Wise[1,2,3,4]

Abstract

Heterochromatic knobs are cytologically-condensed chromosomal regions. However, the structural composition of knobs is not well understood. Here, we report the cloning of KL1HS1, a knob-like sequence from barley, *Hordeum vulgare* L. KL1HS1 consists of seventeen families of diverse tandem repeats. DNA is hypermethylated and recombination is suppressed in the region. KL1HS1 is gene-rich and genes within it are transcriptionally active. KL1HS1 combines the structural characteristics of the two cloned knobs from *Arabidopsis*. It is comprised of a gene-poor core with diverse tandem repeats, a gene-dense island resulting from a 40-kb tandem duplication, and a nested transposable element complex. A positional shift of KL1HS1 was observed and this shifting resulted from a bidirectional expansion of the sequence. Extensive arrays of various tandem duplications in the region are postulated to arise from DNA replication slippage mechanism in combination with a hairpin structure of the knob-associated repeats.

---

[1] Interdepartmental Genetics Program; [2]Department of Plant Pathology, Iowa State University, Ames, IA 50011-1020

[3] Corn Insects and Crop Genetics Research, USDA-ARS, Iowa State University, Ames, IA 50011-1020

[4] Author for correspondence

Introduction

Eukaryotic chromosomes have two distinct forms, lightly stained euchromatin and heavily stained heterochromatin with 4',6-diamidino-2-phenylindole (DAPI). While euchromatin is gene-rich, constitutive heterochromatin, such as knobs, centromeres or telomeres, is gene-poor and consists mainly of tandem repeats or satellite DNA (Yunis and Yasmineh 1976). Heterochomatic regions have been reported to suppress meiotic recombination and gene expression, to delay replication, and to be involved in homologous chromosome segregation (CSHL/WUGSC/PEB *Arabidopsis* Sequencing Consortium, 2000; Fransz et al. 2000). Heterochromatic knobs are comprised mainly of tandem repeats and transposable elements (Ananiev et al. 1998b). In maize, 180- or 350-bp repeats are the major components of knobs in addition to recently inserted retrotransposons. For the 180-bp repeat, there are at least 100 loci in each chromosome (Buckler et al. 1999). These tandem repeats make up to eight percent of the maize genome (Ananiev et al. 1998b). In total, thirty-four visible knobs have been identified among the maize 10 chromosomes. Most of these knobs are telomeric (Buckler et al. 1999). Due to their polymorphic nature, knobs have been used as cytogenetic markers for breeding and phylogenetic analysis. In *Arabidopsis*, two heterochromatic knobs, hk4S and hk5L, have been reported recently (Fransz et al. 2000; Tabata et al. 2000).

Due to the length of chromosomes, it has been difficult to identify knobs in barley by using FISH (fluorescence in situ hybridization) analysis at the pachytene stage, where knobs were first described in maize (Tsuchiya et al., 1984; Carlson 1988; Hang 1994). Here we report molecular cloning of KL1HS1, a barley heterochromatic knob-like sequence, adjacent to the *Mla* powdery mildew resistance locus. Contrary to the generally-held beliefs that knobs are the gene-poor heterochromatic regions, KL1HS1 is gene-rich. KL1HS1 combines

the structural characteristics of the two *Arabidopsis* knobs with various tandem repeats, a nested transposable element complex, and a large fragment duplication. A unique feature of KL1HS1 is that it has undergone a positional shift during evolution of the barley genome.

Results

Various small tandem sequence repeats (Rpt) and simple sequence repeats (SSR) adjacent to the barley *Mla* locus

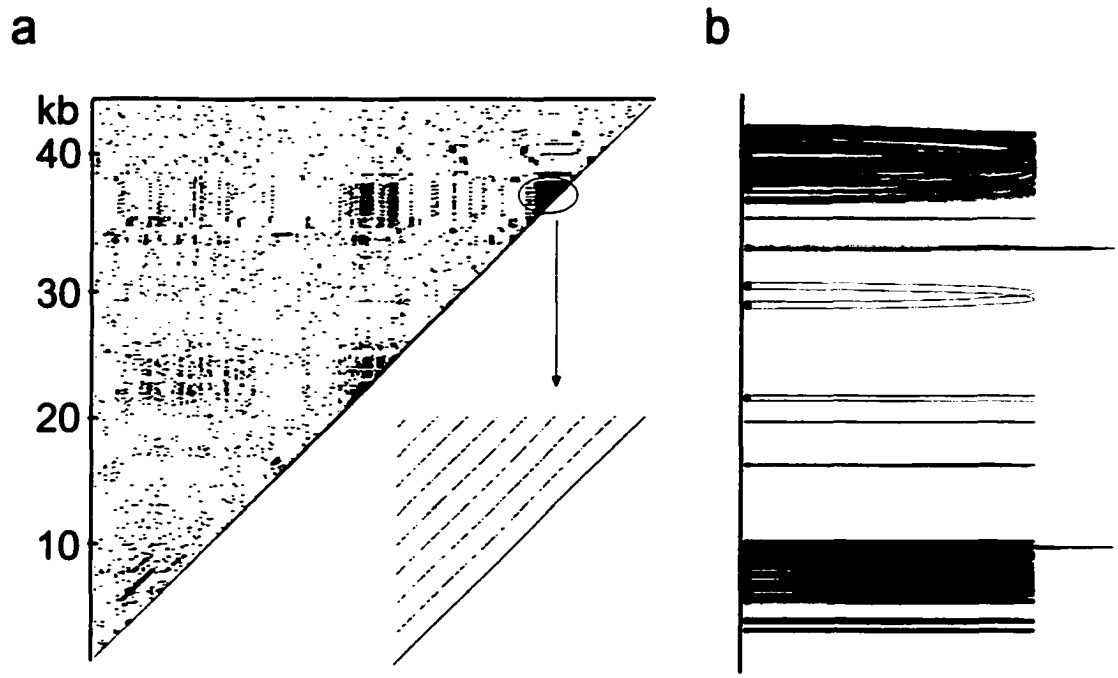
In large-genome plants, individual genes or gene clusters are often separated by large blocks of transposable elements or nested retrotransposon complexes (Moore, 2000). Recently, we reported the gene composition of a 261-kb region spanning the barley *Mla* locus (Wei et al. 2001). The *Mla* region contains three gene-rich islands interspersed with two nested retrotransposon complexes and a gene-poor region. The density is 4.6 kb per gene within the gene islands. However, in the region between genes *Bdi* and *RGH3b*, the gene density is only 16 kb per gene (Wei et al. 2001).

To conduct a detailed analysis of the 45-kb gene-poor region for repetitive sequences and duplication events, we utilized a self-comparison DotPlot program (Fig. 1a). This program facilitated the identification of a 2.5-kb tandem duplication event (straight lines in the lower left of Fig. 1a), several satellite DNA sequences (dense dots in the middle of Fig. 1a), and a block of nine tandem repeats (dense lines in the upper right of Fig. 1a and enlarged region). The position of these repeats could be easily visualized in the output of the analysis by a Miropeat program, which cross-connects all the related repeats (Fig. 1b; Parsons, 1995). In total, there are four SSRs (simple sequence repeat) and seventeen families of tandem repeats, ranging in 2-29 copies (Fig. 1c).

The most significant tandem repeat in the region is an Afa family. The Afa repeat, named after its conserved *Afa* I restriction site, was first identified in *Aegilops squarrosa* L. and was then found to be widespread in the Triticeae (Nagaki et al. 1995). A 24-bp inverted repeat was identified within the Afa sequence. A barley EST BLAST search indicated that this Afa repeat is actively transcribed. These data fully support that the Afa repeat could be a MITE element, which has a terminal inverted repeat, usually inserted into the 3'- or 5'-end of a functional gene and is transcriptionally active (Bureau and Wessler, 1994). Fluorescence in situ hybridization (FISH) analysis has demonstrated that the Afa repeat is distributed in subtelomeric and interstitial regions of chromosomes (Tsujimoto et al. 1997; Nagaki et al. 1999). This is consistent with the location of the *Mla* locus, which is at the telomeric end of barley chromosome 1HS. FISH analysis in pachytene stage can not be easily performed in barley due to the extended length of the chromosomes (Tsuchiya et al., 1984; Hang 1994). Thus, the association of the Afa repeat with a heterchromatic knob has not been determined in barley, as other knob-associated tandem repeats in maize (Chen et al. 2000) and in *Arabidopsis* (Fransz et al. 2000; Tabata et al. 2000). Here, we use molecular and recombination analyses (shown below) to suggest that the Afa repeat is associated with a heterochromatic knob-like sequence.

When we examined the entire 261-kb region, a 40-kb tandem duplication was found directly after repeat Rpt16 (Fig 2a). The 40-kb duplication is gene-rich, comprising seven predicted genes. A pattern of bi-directional expansion of this knob-associated core region can be observed. Flanking the 32-kb region between genes *Bdi* and *3b*, the distal side has a 2.5-kb duplication of gene *g5a/b*, while the proximal side shows the 40-kb duplication with seven genes.

Fig. 1. Barley KL1HS1 containing various kinds of SSR and tandem repeat. This sequence analysis showed a 45-kb region between nucleotide 20,000 to 65,000 of the sequence in the barley Morex *Mla* locus. **a)** Self-comparison of the sequence with 45°-angle lines showing duplication, clustering of lines indicating mutiple tandem duplication. The interval between two lines shows the size of a single repeat element. The clustering of dots suggests a simple sequence repeat region. **b)** Miropeat analysis shows the region of duplication and the sequence relationship among repeats. **c)** A schematic position of each repeat element.


a
b
kb
40
30
20
10

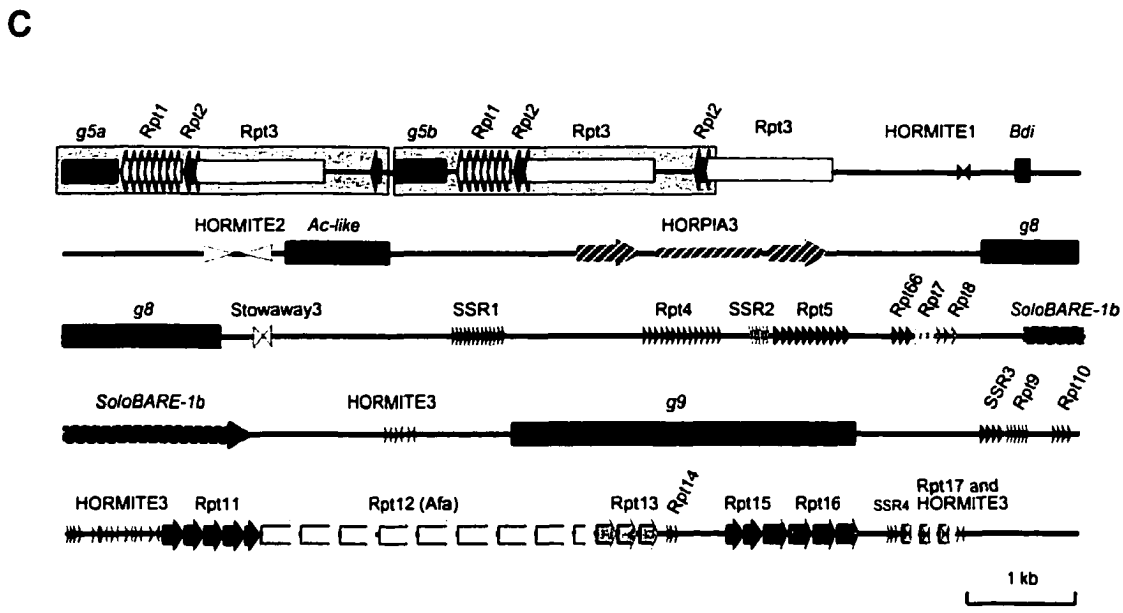
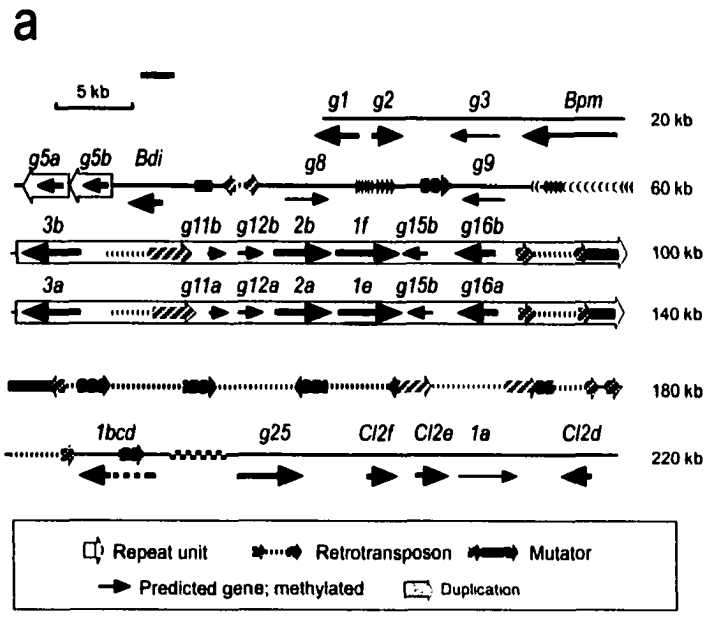
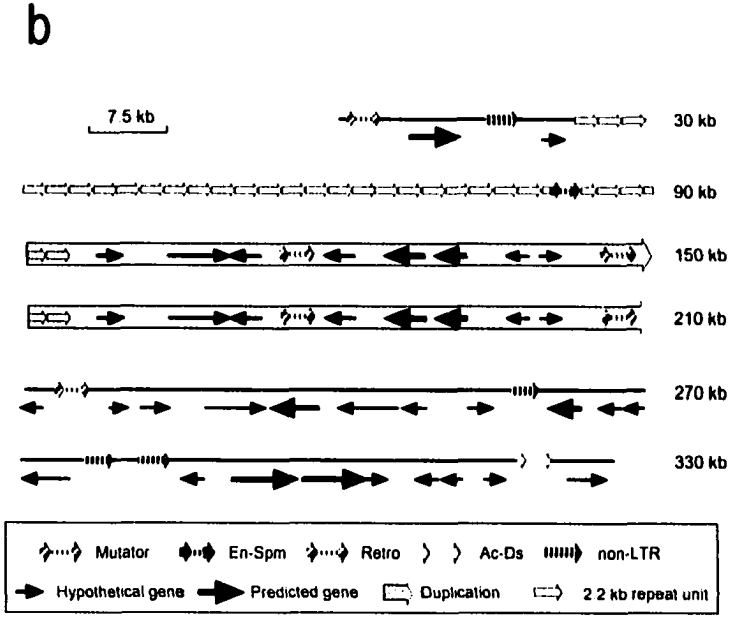

c
g5a
Rpt1
Rpt2
Rpt3
g5b
Rpt1
Rpt2
Rpt3
Rpt2
Rpt3
HORMITE1
Bdi
HORMITE2
Ac-like
HORPIA3
g8
g8
Stowaway3
SSR1
Rpt4
SSR2
Rpt5
Rpt6b
Rpt7
Rpt8
SoloBARE-1b
SoloBARE-1b
HORMITE3
g9
SSR3
Rpt9
Rpt10
HORMITE3
Rpt11
Rpt12 (Afa)
Rpt13
Rpt14
Rpt15
Rpt16
SSR4
Rpt17 and HORMITE3
1 kb

Fig. 2. Parallel structural comparison of barley KL1HS1 and *Arabidopsis* knob hk5L. The filled rectangle with arrow showed duplication. **a)** Overall patterns of barley knob-like sequence KL1HS1. Lines with arrow: predicted genes; red lines: unmethylated genes, navy lines: unmethylated genes, lines with spring green: partial methylated; big arrows: express genes, small arrows: unexpressed genes. **b)** *Arabidopsis* knob hk5L. All genes in this knob are untested for the methylation pattern. Lines with arrow: predicted genes; thick red lines: expressed genes, thin black lines with arrow: unexpressed genes.


a
5 kb
g1 g2 g3 Bpm
20 kb
g5a g5b Bdi g8 g9
60 kb
3b g11b g12b 2b 1f g15b g16b
100 kb
3a g11a g12a 2a 1e g15b g16a
140 kb
180 kb
1bcd g25 Cl2f Cl2e 1a Cl2d
220 kb
Repeat unit
Retrotransposon
Mutator
Predicted gene; methylated
Duplication


b
7.5 kb
30 kb
90 kb
150 kb
210 kb
270 kb
330 kb
Mutator
En-Spm
Retro
Ac-Ds
non-LTR
Hypothetical gene
Predicted gene
Duplication
2.2 kb repeat unit

DNA methylation

DNA is always methylated in heterochromatic regions (Klein and Costa, 1997). To identify the methylation pattern of the 261-kb *Mla* region, we used a PCR-based method. The isorestriction enzymes *Hpa* II and *Msp* I were chosen for the analysis due to their sensitivity to differential methylation. *Msp* I is sensitive only to the methylation of the external C in the restriction site C/CGG while *Hpa* II is sensitive to the methylation of either of the two Cs. If the external C is methylated, a primer pair can amplify a fragment from the template DNA digested with either *Hpa* II or *Msp* I. If only the internal C is methylated and the external C is not methylated, a PCR product can be amplified from the *Hpa* II-cut template, but not from the *Msp* I-cut DNA template. Fig. 3a through 3d illustrates the amplification of DNA digested with *Hpa* II or *Msp* I in four barley accessions. Cv. Morex is the origin of this sequence. Morex and Steptoe are parents of a common barley mapping population (Kleinhofs et al. 1993) whereas C.I. 16151 and C.I. 16155 are parents of a high resolution mapping population spanning the *Mla* region (Wei et al. 1999). In summary, the genes between *g2* and *1bcd* are methylated while genes proximal and distal to this region are not methylated except for *CI2e* (Fig. 2a and Table 1). The sequence between *g5b* and *1bcd* is hypermethylated except *RGH3a/b*. This hypermethylated region includes promoters, 5'- or 3'-termini, coding region of a gene or in an intergenic region. Genes *Bpm*, *g5a/b*, *RGH3a/b* and *CI2e* are methylated only at the internal C, thus not hypermethylated.

Recombination suppression

Heterochromatin normally suppresses meiotic recombination (Grewal and Klar, 1997; CSHL/WUGSC/PEB *Arabidopsis* Sequencing Consortium, 2000). In our previous study (Wei et al. 1999), no recombination events among 3,600 gametes were detected in an over-200-kb region. The recombination rate is over ten folds lower than in the adjacent region.

Fig. 3. Expression and methylation analyses of genes in barley *Mla* locus. Primers, amplification conditions and the summary results are in Table 1. **a), b), c), and d)** show the different patterns of DNA methylation in PCR analysis. Positive amplification indicates the DNA is methylated, can not be cut by the restriction enzyme and have a PCR product. No amplification shows DNA is not methylated, can be digested and should not have a PCR product. *Hpa* II, *Msp* I and Genomic indicate barley DNA was cut with the respective enzyme while genomic shows uncut genomic DNA as a positive PCR control. a), b), c), and d) represented genes *Bpm*, *1bcd*, *g9* and *g2*, respectively. See Table 1 for full name of these genes. **e)** RT-PCR analysis of transcription expression of predicted genes. Above the panel show some components in each one-step RT-PCR analysis. RNA or DNA indicates to use RNA or genomic DNA as a template for the reaction. RT: reverse transcriptase. In the analysis of each gene, the first lane shows a real reaction of RT-PCR. The second and the third lanes represent a DNA-free RNA negative control with no reverse transcriptase added and a positive PCR amplification.

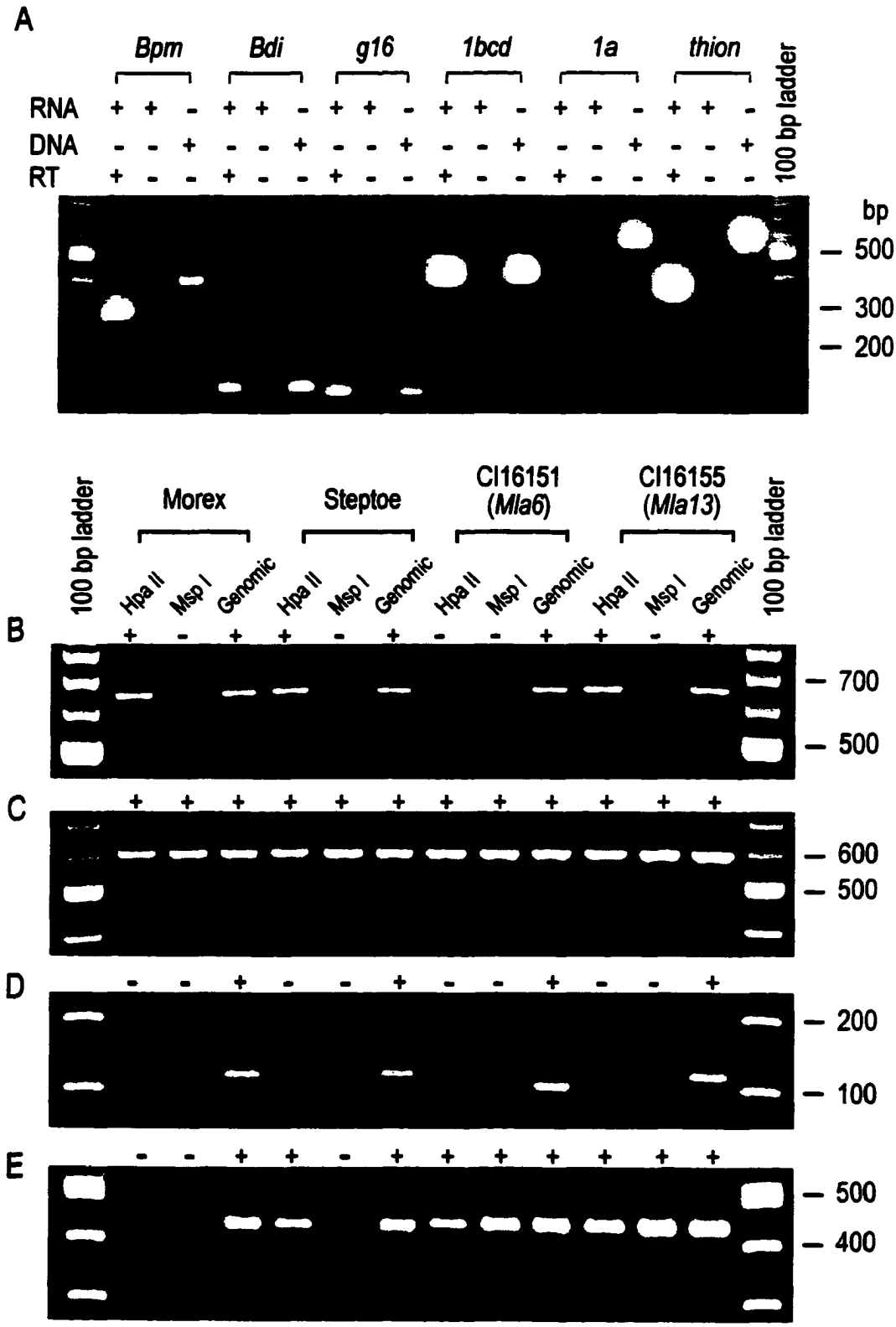

A
Bpm
Bdi
g16
1bcd
1a
thion
RNA
DNA
RT
100 bp ladder
bp
500
300
200
B
100 bp ladder
Morex
Steptoe
CI16151 (Mla6)
CI16155 (Mla13)
Hpa II
Msp I
Genomic
100 bp ladder
700
500
C
600
500
D
200
100
E
500
400

Table 1. Methylation and expression tests for predicted genes at the barley knob1HS1 region

Gene Abbr.	Gene	Primer Pair	Anneal. Temp.	Size (bp)	Methylation Test[a]				Transcript. Expression
					Morex	Steptoe	*Mla6*	*Mla13*	
g1	*71IN16.16*	CGTAGATAGACAACGGAATCGAGG GGGCGGGGTTCAACATATTG	58	271	-	+	-	-	+
g2	*71IN16.15*	CGGACGAACAGTTGGTGTGC ACCAGGAGGAAGGAACACGC	62	215	-	-	-	-	+
g3	*71IN16.14*	CGATCACAGTCACCGAGCAATC TGAAGATAGCCAACGCCCTG	58	138	++	++	++	++	-
Bpm	*Bpm*^{b}	ATAAGCCGCCTTGCATAATACG AACTACCACTTCACACCTTGCAGG	55	625	+	+	+	+	+
g5a&b	*Bmt*^{b}	TTCTAGCAAGCATCTAAAGGTTTCG ACATCTGTTAGTCGGGCAAGAGAC	54	576	+	+	+	+	-
Bdi	*Bdi*	AGGGCATCTTCAATGGTCCTATG CGGAGCTTGTGCGTGAACC	51	136	++	++	++	++	+
g8	*71IN16.9*	CGACATTGTCCTGCTCTGTTGC CGGTTTGAATTGGACGAGCTTG	58	134	++	++	NP	++	+
g9	*71IN16.8*	TGTTGACTCCTTGATTCCATCCAC CTGCTAGATAAAGGCATCGTACCTG	57	582	++	++	++	++	+
3a&3b	*RGH3*	GGTGTGTGATTTCGATGCC CAGGAGCCTGCACCGTCT	54	146	+	+	+	NP	+
g11a&b	*80H14.15*	ACCCTGCTGCCGATCTACTGATAG ATCGTTCCTCTTCCTCGTCGTC	59	157	++	++	++	++	-

Table 1. (continued)

Gene Abbr.	Gene	Primer Pair	Anneal. Temp.	Size (bp)	Methylation Test[a]				Transcript. Expression
					Morex	Steptoe	*Mla6*	*Mla13*	
g12a&b	*80H14.14*[b]	GAGCACACTGGATTGTTGAAATGG CACCCATGCCTTCCATAGTAGC	58	374	++	++	++	++	-
2a&b	*RGH2*	ACCCTCGCCAGACAAGTTTACC GAAAAGCAGCTTATGCACGTCG	57	165	++	NP	++	NP	+
1e&f	*RGH1e/f*	CATGGTTAGCCTGATCTCCAAGC TGGAATTGCTCCGTGATTTCC	57	372	++	NP	++	NP	+
g15a&b	*80H14.11*	TCACACCATACGCCGCAAC AGGTGCTCACCGCAACGTC	64	378	++	++	++	++	-
*g16a*b*	*80H14.10*	CCTTTGTCGTTAGGATCGCATTC GTTGGTAGGGTTATTGTTCATGCAC	56	131	++	++	++	++	+
1bcd	*RGH1bcd*	GAGTAATTGTCGCCGTTTGTCC GCAACACACCAACTAGAGGAAACAC	54	434	-	+	+	++	+
g25	*80H14.8*[b]	TCGGTGACGGTGTGACAATTC CTTGATCTTCCCCTTGCTTGC	58	493	-	NP	-	-	+
Cl2f	*Cl2f*	GGACCTGGGCAATGTTGTCG CAAAGGAAGCAAGCGTAACAAGG	58	264	-	-	-	+	+
Cl2e	*Cl2e*	GCAACCCAACTAGCCAACGTG GCCAAGGAGATCATTCTCAAGGAC	59	150	+	+	+	++	+
1a	*RGH1a*	TTGGGAAAAATAGCAGCCTGC ATGCAAGGGGTCTCTCATTATCC	56	572	-	++	-	-	-

Table 1. (continued)

Gene Abbr.	Gene	Primer Pair	Aneal. Temp.	Size (bp)	Methylation Test[a]				Transcript. Expression
					Morex	Steptoe	*Mla6*	*Mla13*	
Cl2d	*Cl2d*	GCTGACCCAAAAGCACCCG CTAGCCGATGTGAGGAGTGGTC	59	216	-	-	-	-	+

[a]Methylation test: +: only Hpa II sensitive, internal C methylated; ++: both Hpa II and Msp I sensitive, external or both Cs methylated; - :not methylated; NP: No PCR product amplified from unrestricted genomic DNA.

[b]These primers are for methylation test only; for RT-PCR, primers GAGCACACTGGATTGTTGAAATGG and GGAGCTTGATAGGTGTTGGCAAG are used at 55 °C to get 169 bp *80H14.14* product; primers AGAAAAGTTCCAGAAGATG and AAAAGACTATTTCGATTCC at 44 °C for 117 bp *Bmt* product; primers ACTTCTTCAGTGTGTTCAGGTGAGC and GCAAGCCACATGAGAGAACTGC at 55 °C for 387 bp *Bpm* product; primers GCTACTGGAGCCATCACATGAAG and CTTGATCTTCCCCTTGCTTGC at 58 °C for 209 bp *80H14.8* product.

We had postulated that this may be the result of a lack of pairing and subsequent strand exchange between homologous regions in the C.I. 16151 and C.I. 16155 parents of our mapping cross. Recombination suppression was also observed in the tomato *Mi* (Van Daelen et al. 1993), *Tm2-a* (Ganal et al. 1989) and the lettuce *Dm3* (Chin et al. 2001) resistance loci. The *Mi* and *Tm-2a* loci are physically adjacent to the pericentromeric heterochromatin where recombination is normally suppressed. In lettuce *Dm3* downy mildew resistance locus, there is a region with 18 fold lower in recombination rate (Chin et al. 2001). Because both *Mla* and *Dm3* are telomeric, with our finding that *Mla* gene is next to a heterochromatic knob, it would be interesting to know whether there is a heterochromatin region close to *Dm3* gene. There is a report in the *Dm3* locus that two spontaneous mutants with over 500-kb to 1Mb segment deletion have 3- and 3.5-fold increases in recombination rates, respectively (Chin et al. 2001). It could be explained by deletion of a possible heterochromatic knob in the region.

Active gene transcription

Gene transcription is usually repressed in heterochromatic regions. To investigate the gene expression pattern of the knob-associated region in the 261-kb barley *Mla* locus, EST BLAST searches, cDNA screening and RT-PCR analyses were conducted. BLASTn searches of nearly 69,000 barley ESTs showed that genes *Bpm*, *Bdi*, *g25*, *CI2f*, *CI2d*, and *CI2c* are transcribed (Fig. 2a; Table 1). It is interesting to note the differential gene expression of the six-membered *CI2* (chymotrypsin inhibitor 2) gene family. While the *CI2f* gene was detected only among ESTs from 5-45 DAP (days after pollination) spike libraries, it was not detected in the seedling leaf, root, and shoot libraries. The *CI2c* and *CI2d* genes were detected in seedling root, shoot and leaf, but not detected in the spike.

Although transcripts of *Mla-RGH1*, *RGH2*, and *RGH3* family members were not detected via BLASTn searches, two hundred *Mla-RGH1* family clones were identified from a total of 1.2 million plaques in a cDNA library from barley accession C.I. 16155 (*Mla13*). Similarly,

nineteen *RGH1* family clones were identified in a total of 0.6 million plaques from a cDNA library of barley accession C.I. 16151 (*Mla6*) (Halterman et al. 2001). No clones were retrieved with probes for either the *RGH2* or *RGH3* family in either library screen. Our previous study demonstrated that *RGH1bcd* is the Morex allele of *Mla6* and *Mla1* (Wei et al. 2001). From this cDNA analysis for the three *RGH* families, we treated *RGH1bcd* as being actively transcribed (Fig. 1a; Table 1).

Limited by the sample sizes in the barley EST database and in cDNA screening, some expressed genes may not be detected by these two analyses. To further increase the screening sensitivity for gene transcription, a RT-PCR analysis was employed. Genes *g2*, *Bpm*, *Bdi*, *g8*, *g9*, *3a/b*, *2a/b*, *1e/f*, *g16a/b*, *1bcd*, *CI2e*, and *CI2d* were found to be transcribed in this assay. The combined results of these three analyses were summarized in Table 1 and Fig. 2a. From the analysis of gene transcription with EST database searches, cDNA library screening, and RT-PCR amplification, we concluded that most genes in the region are transcribed, no matter where they are within KL1HS1 or flanking this region.

We did not detect a correlation between methylation and transcription in the 261-kb region. Genes *g8*, *g9*, *3a/b*, *2a/b*, *1e/f*, and *g16a/b* are hypermethylated, but are transcribed. Genes *Bpm*, *RGH3a/b* and *CIe* are transcribed while they are methylated, but not hypermethylated. Genes *g1* and *RGH1a* are not transcribed even though they are not methylated. Our results are in contrast to the long observed high correlation between methylation and gene silencing (Paszkowski and Whitham, 2001).

Discussion

Barley KL1HS1 combines the structural characteristics of the two cloned *Arabidopsis* knobs

The region between genes *Bmta* and *1bcd* harbors the characteristics of a heterochromatin knob, such as arrays of tandem repeat, hypermethylation, and meiotic recombination. These characteristics suggested a possible heterochromatic knob. We designated the region as knob-like sequence in the barley chromosome 1HS as KL1HS1, the first knob-like sequence in chromosome 1HS. A physical boundary of KL1HS1 can postulated by combining the tandem repeats and a pattern of methylation. Between genes *Bmta* and *1bcd*, the region is composed of 17 families of tandem repeat and are hypermethylated except *RGH3a/b*, which are methylated, not hypermethylated. This region comprises three structural complexes; a 32-kb gene-poor core, an 40-kb tandem duplication of gene-rich region and a 52-kb nested transposable element complex (Fig. 2a).

The barley KL1HS1 is structurally similar to the two cloned knobs, hk5L and hk4S, from *Arabidopsis*. The knob hk5L contains 66-kb of small tandem duplications from 30 copies of a 2.2-kb repeat in addition to a nearly 60-kb large tandem duplication (Tabata et al. 2000; Fig. 2b). The large tandem duplications in both barley and *Arabidopsis* are gene-rich and contain a low percentage of transposable elements. The other *Arabidopsis* knob, hk4S, has 44-kb of small tandem duplications derived from 22.5 copies of a 1,950-bp repeat in addition to the flanking 139-kb transposons and retrotransposons (CSHL/WUGSC/PEB *Arabidopsis* sequencing Consortium, 2000). Overall, all the three knobs have tandem arrangement of small repeat elements. The barley KL1HS1 is highly similar to *Arabidopsis* hk5L in the large duplication events while it resembles *Arabidopsis* hk4S in the transposable element complexes. Therefore, we can conclude that barley KL1HS1 combines the structural

characteristics of the two *Arabidopsis* heterochromatic knobs. It is interesting to notice that a knob seems to be at least 160 kb with knobs *Arabidopsis* hk4S 180 kb, hk5L 190 kb and barley KL1HS1 160 kb. The maize knobs could be up to several megabase (Ananiev et al., 1998a).

Dynamics of KL1HS1, including heterochromatin expansion, shifting and positional activation

The unusual feature of KL1HS1 is that it contains many different sequence repeats, including four different SSRs and seventeen distinct tandem repeats. In the two cloned *Arabidopsis* knobs, each contains only one family of repeat, the 1,950-bp repeat in knob hk4S and the 2,200-bp repeat in knob hk5L. The duplication of tandem repeat in *Arabidopsis* hk4S has been suggested from preferential integration of the transposon-like 1,950-bp repeat. This preferential insertion was proposed from the chromatin condensation by the existence of a chromatin-binding site in the tandem repeat (CSHL/WUGSC/PEB *Arabidopsis* sequencing Consortium, 2000). Up to now, there is no evidence of such chromatin-binding site (CBS). If there is a CBS, the site should be be conserved. However, a pileup of the consensus sequences of all available knob-associated repeats from GenBank did not reveal such a conserved chromatin-binding site. These knob-associated repeats include the maize 180- and 350-bp repeats, the barley Afa repeat, the *Arabidopsis* hk5L-associated 2,200-bp repeat and the 160-bp consensus sequence of hk4S-associated 1,950-bp with its related pericentromeric 1-kb repeat. Interestingly, in DNA secondary structure similarity analysis, we did find similar structures between the barley 340-bp Afa repeat and the maize 350-bp repeat and between the maize 180-bp repeat and *Arabidopsis* hk4S 1,950-bp repeat related 160-bp sequence (data not shown). The AT-rich feature of these repeats, from 58% to 73%, could explain the similar secondary structure and also suggests the

importance in these structures. These structural similarities perhaps indicate that the structure, not the sequence, is more important to the supposed chromatin-binding protein.

The intact nature of each knob-associated repeat and the diversity of the tandem repeats, from 20 bp to 40 kb, strongly implies that the transposon insertion hypothesis for the origin of tandem repeat is not the case. If a tandem duplication resulted from transposon insertion, the host repeat would be disrupted. However, no such disruption was observed. The two large tandem duplication in both knobs of barley KL1HS1 and *Arabidopsis* hk5L is not a transposon, but contains multiple genes. Instead of transposon insertion, these duplications are more likely to result from a mechanism of DNA replication slippage (Fig. 4; Weston and Berg, 1991; Trinh and Sinden, 1991). The palindromic or quasi-palindromic sequences are essential for this slippage mechanism to form complex secondary structures in single-stranded DNA. The evolution of satellite DNA has been thought to result from DNA replication slippage (Awadalla and Ritland, 1997). In the KL1HS1 region, the palindromic sequences are rich and are characterized as MITE elements. All fifteen MITE elements can form hairpin structures. Most of these MITEs are positioned directly upstream of the tandem duplication. For example, repeat Rpt17 could be explained with a DNA slippage mechanism. Most of other Rpt elements also have quasi-palindromic sequence or can form stable secondary structure and thus enhance the DNA replication slippage.

Sequence similarity analysis of these different tandem duplications in different subregion of KL1HS1 suggests an evolutionary pattern of KL1HS1. By comparing the sequence similarity of each repeat family, we found a functional shifting of KL1HS1. Inside the 2.5-kb duplication unit, there is a family of eight repeats that share about 70% similarity (Rpt1; Fig. 1a). The 2.5-kb tandem repeat shares 75% DNA identity. The low similarity of these repeats, in addition to truncated feature of *g5* gene family, strongly implied that the region originated from an possible ancient knob. On the proximal side, the DNA similarities of

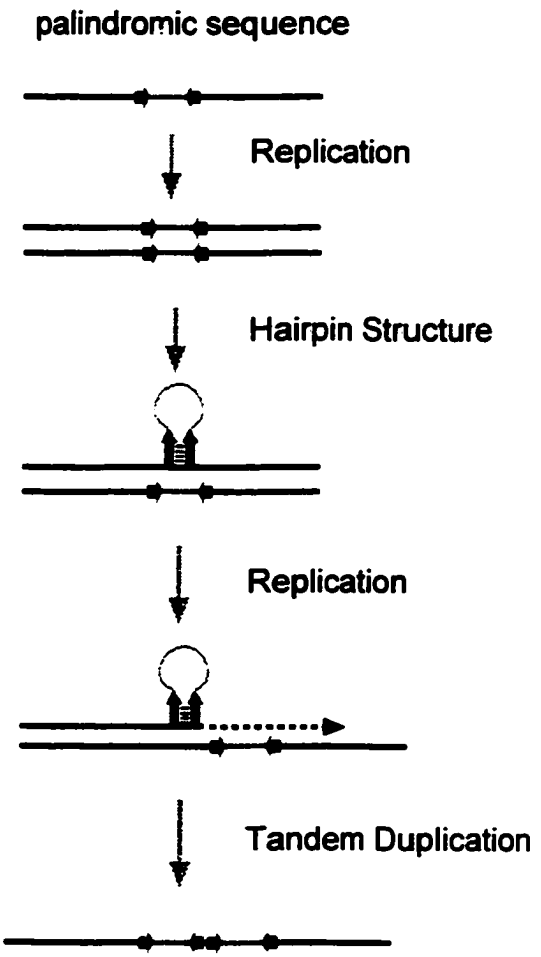


Fig. 4. The propagation of tandem repeat in the genome by a DNA replication slippage mechanism.

Rpt4 family range from 82-89%, whereas Rpt5 (Afa) shares 96.5 to 98.2% identity, Rpt6 92-96% identity, Rpt7 90% identity, Rpt8 from 85-95% identity, and Rpt9 from 99 to 100% identity, and the nearly 40-kb duplication is 99.9% identity. The higher DNA similarities suggest that this side of KL1HS1 is developed more recently. From Fig. 2a, we know genes *g5* and *Bpm* are partially methylated. Combining those genes with the sequence similarities, we can postulate that the distal end of the original KL1HS1 was degenerated and the possible knob shifts to proximal end side by large tandem duplication and nested transposable element insertion. Indeed, the pericentric heterochromatin shift of centromere was observed in *Drosophila* by transposon insertions (Henikoff, 2000).

It is also worth noting that the sequence similarities indicate that the propagation of Afa repeat is coincident with the large scale of retrotransposon insertion in the barley genome in the past three million years (data not shown; Wei et al. 2001). It is possible that the entire genome expansion needs a corresponding knob expansion to keep the genome functional. Under this circumstance, the insertion or duplication of the Afa repeat triggered the most recent 40-kb tandem duplication.

Finally, we propose a scenario of how a heterochromatic knob forms. An AT-rich knob-associated repeat (KAR) inserts itself into a genome region and gets propagated with the DNA replication slippage mechanism. If the insertion site is in a TE (transposable element)-rich region, or becomes a TE-preferred insertion environment after KAR propagation, the regions turn out to be a knob. Alternatively, if the insertion site is gene-rich, it will duplicate the surrounding gene-rich regions to a suitable size to be a heterochromatic knob.

Materials and Methods

Sequence analysis: Repeat sequences were identified by web-based Large Dot Plots program (http://alces.med.umn.edu/rawdot.html) in combination with Miropeat program (Parsons, 1995). Sequence similarity analysis was conducted by the GCG-GAP program (Wisconsin Package for sequence analysis; Oxford Molecular, Madison, WI). The GCG-Pileup and Prettybox programs were used for multiple sequence comparison and visualization. PAUP4.0 software (http://www.lms.si.edu/PAUP/) was used for phylogenetic analysis of tandem repeats. MacVector v6.0 (Oxford Molecular) program was adopted to design PCR primers and to map restriction enzyme sites. EST (expressed sequence tag) searcher were performed by the NCBI BLASTn program.

Methylation test: A PCR-based method was employed for the methylation analysis. While restriction enzyme *Msp* I is sensitive to the methylation of only the external C nucleotide of the C/CGG restriction site, the isorestriction enzyme *Hpa* II is sensitive to the methylation of either C. Barley genomic DNA was first digested with either *Msp* I or *Hpa* II. The restricted DNA was then used as template for PCR amplification. PCR primers were designed such that a product can not be amplified from the completely restricted DNA while a product should be amplified from the methylated template. Most of the primers are from the coding regions of genes. In the case of no restriction site in the coding region, primers designed in regions that are 2-kb up- or down-stream of the respective gene. DNA was digested with *Hpa* II or *Msp* I (10 units enzyme / μg DNA; New England Biolabs, Beverly, MA) at 37°C for 10 hr, and the enzymes were then inactivated at 65°C for 20 min. PCR [Taq DNA polymerase from Life Technologies (GIBCO BRL), Gaithersburg, MD] was performed to amplify the sequence. PCR primers and annealing temperatures are shown in Table 1.

Reverse Transcription PCR (RT-PCR) analysis: Total RNA was isolated from 7-day old barley seedling as instructed in RNeasy Plant Mini kit (QIAGEN, Valencia, CA). The isolated total RNA was further digested with DNase I (Ambion, Austin, TX) and purified by the clean-up application of the above QIAGEN kit. RT-PCR primer pairs are the same as those used in methylation test except those from genes which do not have a *Msp* I or *Hpa* II restriction site in the coding region. In the latter situation, new RT-PCR primer pairs were designed according to the predicted coding region. The sequences of PCR primers and their annealing temperatures are shown in Table 1. Because most primer pairs are within an exon of the respective genes, strict controls were performed to make sure that the total RNA was DNA-free. The DNase cleaned total RNA was first tested with primer pairs from genes whose transcription was confirmed via a barley EST database search. Genomic DNA was used as a positive PCR control. In the RT-PCT analysis of each gene, a negative control with only Taq polymerase but no reverse transcriptase added and a positive genomic DNA control were performed. RT-PCR was conducted according to Superscript One-step RT-PCR kit (Gibco-BRL) with 0.5 μg total RNA per reaction. The amplified DNA fragments were resolved in 2% LE agarose gel (FMC Bioproducts, Rockland, ME) with ethidium bromide.

Acknowledgements

The authors thank L. Mao and R. A. Wing for assistance of Miropeat analysis. The authors also thank S. Whitham (Dept. of Plant Pathology, ISU) for helpful advice for the RT-PCR analysis and discussion on gene silencing. This research was supported in part by USDA-NRI/CGP grants 98-35300-6169 & 00-35300-9213 to RPW. Joint contribution of the Corn Insects & Crop Genetics Research Unit, USDA-Agricultural Research Service and the Iowa Agriculture and Home Economics Experiment Station. This is Journal Paper No. J-

xxxxx of the Iowa Agricultural and Home Economics Experiment Station, Ames, IA., Project No. 3368, and supported by Hatch Act and State of Iowa funds.

References

Ananiev, E.V., Phillips, R.L., and Rines, H.W. (1998a) A knob-associated tandem repeat in maize capable of forming fold-back DNA segments: are chromosome knobs megatransposons? Proc. Natl. Acad. Sci. USA. 95:10785-10790.

Ananiev, E.V., Phillips, R.L., and Rines, H.W. (1998b) Complex structure of knob DNA on maize chromosome 9. Retrotransposon invasion into heterochromatin. Genetics 149:2025-2037.

Awadalla, P., and Ritland, K. (1997) Microsatellite variation and evolution in the Mimulus guttatus species complex with contrasting mating systems. Mol. Biol. Evol. 14:1023-1034.

Buckler, E.S. 4th, Phelps-Durr, T.L., Buckler, C.S., Dawe, R.K., Doebley, J.F., and Holtsford, T.P. (1999) Meiotic drive of chromosomal knobs reshaped the maize genome. Genetics 153:415-426.

Bureau, T.E., and Wessler, S.R. (1994). *Stowaway*: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell 6:907-16.

Carlson, W.R. (1988) The cytogenetics of corn. In: Sprague, G.F. and Dudley, J.W. (edr) Corn and Corn Improvement. American Society of Agronomy, Madison, Wis., pp259-344.

Chen, C.C., Chen, C.M., Hsu, F.C., Wang, C.J., Yang, J.T., and Kao, Y.Y. (2000) The pachytene chromosomes of maize as revealed by fluorescence in situ hybridization with repetitive DNA sequences. Theor. Appl. Genet. 101:30-36.

Chin, D.B., Arroyo-Garcia, R., Ochoa, O.E., Kesseli R.V., Lavelle, D.O., and Michelmore, R.W. (2001) Recombination and spontaneous mutation at the major cluster of resistance genes in lettuce (*Lactuca sativa*). Genetics 157: 831-849.

CSHL/WUGSC/PEB *Arabidopsis* sequencing Consortium, (2000) The complete sequence of a heterochromatic island from a higher eukaryote. Cell 100:377-386.

Fransz, P.F., Armstrong, S., de Jong, J.H., Parnell, L.D., van Drunen, C., Dean, C., Zabel, P., Bisseling, T., and Jones, G.H. (2000) Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region. Cell 100:367-376.

Ganal, M.W., Young, N., and Tanksley, S.D. (1989) Pulsed field gel electrophoresis and physical mapping of the large DNA fragments in the *Tm-2a* region of chromosome 9 in tomato. Mol. Gen. Genet. 215:395-400.

Grewal, S.I., and Klar, A.J. (1997) A recombinationally repressed region between mat2 and mat3 loci shares homology to centromeric repeats and regulates directionality of mating-type switching in fission yeast. Genetics 146:1221-1238.

Halterman, D., Zhou, F., Wei, F., Wise, R.P., Schulze-Lefert, P. (2001) The MLA6 coiled-coil, NBS-LRR protein confers AvrMla6-dependent resistance specificity to *Blumeria graminis* f. sp. *hordei* in barley and wheat. Plant J. 25:335-348.

Hang, A. (1994) Use of an acrocentric chromosome in a genetic and linkage study of barley chromosome 4. Cytologia 59:187-193.

Henikoff, S. (2000) Heterochromatin function in complex genomes. Bioch. Bioph. Acta. 1470:O1-O8.

Klein, C.B., and Costa, M. (1997) DNA methylation, heterochromatin and epigenetic carcinogens. Mutation Research 386 :163-180.

Kleinhofs, A., Kilian, A., Saghai-Maroof, M.A., Biyashev, R.M., Hayes, P., Chen, F.Q., Lapitan, N., Fenwick, A., Blake, T.K., Kanazin, V. (1993) A molecular, isozyme and morphological map of the barley (Hordeum vulgare) genome. Theor. Appl. Genet. 86:705-712.

Moore, G. (2000) Cereal chromosome structure, evolution, and pairing. Annu. Rev. Plant Physiol. Plant Mol. Biolol. 51:195-223.

Nagaki, K., Tsujimoto, H., Isono, K., and Sasakuma, T. (1995) Molecular characterization of a tandem repeat, Afa family, and its distribution among Triticeae. Genome 38:479-486.

Parsons, J.D. (1995) Miropeats: graphical DNA sequence comparisons, Comput. Applic. Biosci. 11:615-619.

Paszkowski, J., and Whitham, S.A. (2001) Gene silencing and DNA methylation processes. Curr. Opin. Plant Biol. 4:123-129.

Tabata, S., Kaneko, T., Nakamura, Y., Kotani, H., Kato, T., et al. (2000) Sequence and analysis of chromosome 5 of the plant *Arabidopsis thaliana*. Nature 408:823-826.

Trinh, T.Q., and Sinden, R.R. (1991) Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *Escherichia coli*. Nature 352:544-547.

Tsuchiya,T., Singh, R. J, Shahla, A., and Hang, A. (1984) Acrotrisomic analysis in linkage mapping in barley (*Hordeum vulgare*). Theor. Appl. Genet. 68: 433-439.

Van Daelen, R.A.J.J., Gerbens, F., Van Rusissen, F., Aarts, J., and Honteleez, J. et al. (1993) Long-range physical maps of two loci (*Aps-1* and *GP79*) flanking the root-knot nematode resistance gene (*Mi*) near the centromere of tomato chromosome 6. Plant Mol. Biol. 23:185-192.

Wei, F., Gobelman-Werner, K., Morroll, S.M., Kurth, J., Mao, L., Wing, R., Leister, D., Schulze-Lefert, P., and Wise, R.P. (1999) The *Mla* (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. Genetics 153:1929-1948.

Wei, F., Wing, R.A., and Wise, R.P. (2001) Function clustering during molecular evolution of a plant defense locus. In preparation.

Weston, H.K, and Berg, D.E. (1991) Limits to the role of palindromy in deletion formation. J. Bacteriology 173:315-318.

Yunis, J.J., and Yasmineh, W.G. (1971) Heterochromatin, satellite DNA and cell function. Science 174:1200-1209.

CHAPTER 5. GENERAL CONCLUSION

Conclusions

Powdery mildew of barley, caused by *Blumeria* (*Erysiphe*) *graminis* f. sp. *hordei*, is a model system for investigating the mechanism of gene-for-gene interaction between large-genome cereals and obligate-fungal pathogens. A large number of loci that confer resistance to this disease are located on the short arm of chromosome 5(1H), such as the *Mla* locus in the telomeric end of 1HS. We used AFLP-, RAPD-, and RFLP-derived markers to saturate the *Mla* region in a high-resolution recombinant population segregating for the (*Mla6* + *Mla14*) and (*Mla13* + *Ml-Ru3*) resistance specificities. These tightly linked genetic markers were used to identify and develop a physical contig of YAC and BAC clones spanning the *Mla* cluster. A low-pass sequencing strategy in combination with BAC-end sequencing revealed three distinct NBS-LRR resistance-gene homologue (*RGH*) families in the contig spanning the *Mla* cluster. Genetic and physical mapping delimited the *Mla*-associated, NBS-LRR gene families to a 240-kb interval. Recombination within the *RGH* families was at least 10-fold less frequent than between markers directly adjacent to the *Mla* cluster.

The 261,265-bp complete sequence of two overlapping BACs spanning the *Mla* locus revealed thirty-two protein-encoding and two $tRNA^{ser}$ genes. Gene dense islands and nested transposon insertions are common features of the region in this large-genome cereal plant. The protein-encoding genes are organized as three gene-rich islands separated by two 40-kb complexes of nested transposable elements and a gene-poor region. Sixteen of these genes are plant-defense related; 12 of these 16 are associated with defense against powdery mildew disease, but function in different signaling pathways. The clustering of those plant defense-related genes is similar to that in the mammalian major histocompatibility complex (MHC;

(Beck and Trowsdale, 2000). Extensive duplication and TE insertion mainly contribute to the genome expansion of the sequence. Evolutionary analysis indicated that the present *Mla* region was developed over 7 million years through several duplication and inversion events in addition to nested TE insertion.

A heterochromatic knob-like sequence, KL1HS1, is present in this barley *Mla* complex. KL1HS1 is hypermethylated and recombination in the region is suppressed. KL1HS1 combines the structural characteristics of the two cloned knobs from *Arabidopsis* (CSHL/WUGSC/PEB *Arabidopsis* sequencing Consortium 2000; Franzs et al. 2000; Tabata et al. 2000). This knob-like sequence is comprised of a gene-poor core with diverse tandem repeats, a gene-dense island resulting from a 40-kb tandem duplication, and a nested transposable element complex. In contrast to other knobs, KL1HS is gene-rich and transcriptionally active. A positional shift was observed and this shifting resulted from a bidirectional expansion of KL1HS1. Extensive arrays of various tandem duplications in the knob are postulated to arise from a DNA replication slippage mechanism in combination with a hairpin structure of the knob-associated repeats.

Recommendation for Future Study

In this study, we found that most of genes in this *Mla* region are functionally connected via different pathways. More surprisingly, a heterochromatic knob-like sequence was uncovered in the region. This knob-like region contributes to the recombination suppression observed in our previous study (Wei et al. 1999; Chapter 2).

Exemplified by the barley *Mla* complex, locus genomics studies the genome structure and function of a gene locus and its flanking region. As a proponent of such a concept, the following research needs to be focused in the future. First of all, FISH (fluorescence in situ

hybridization) analysis is essential to further confirm our molecular data, which suggest a knob in the region. If FISH data support the existence of a knob in the region, it would be significant because the genes in the region are active. Meanwhile, an analysis of transcription expression of genes in the *Arabidopsis* heterchromatic knob hk5L, which is similar in structure to KL1HS1, would also be interesting because two genes in the 60-kb tandem duplication is also expressed in the EST analysis. The second is the *Mla*-mediated signal pathway. After cloning of functional *Mla* genes with assistance of this research (Halterman et al. 2001; Zhou et al. 2001), identification of genes involved in the resistance signaling is a natural next step. Indeed, over forty mutants have been accumulated (R. P. Wise, unpublished). Because the functions of *RGH2* and *RGH3* families are unknown, it would also useful to map them in other species and to find out their linkage relationship with any known resistance gene loci.

Thirdly, a barley *Pum/Mpt5/FBF*-like (*Bpm*) gene was found in the *Mla* region. The products of *Pum/FBF/Mpt5* genes (*Puf* family) have been shown to repress mRNA translation by binding to 3' UTR of the RNAs in *Drosophila*, *C. elegans*, and yeast (Tadauchi, et al. 2001). Three putative *Puf* orthologs have been reported in *Arabidopsis* (Lin et al., 2000). In plants, the function of genes in the *Puf* family is yet untested and it would be interesting to investigate that. Because of the conserved feature of the genes in these two diverse species, they should be functionally important. The study of the function of *Bpm* gene can be done in the model plant *Arabidopsis*, not necessary in barley.

Fourthly, differential expression of a duplicated multiple *CI2* gene family was observed. The six-membered gene family is conserved only in the coding regions, but diverse in other regions, such as 5'-UTR and promoters. The differential expression of a duplicated *myb* gene family has recently been reported in maize (Zhang et al. 2000). To further characterize the observed differential expression, we can replace the coding regions of the *CI2* genes with

a reporter gene and test their expression patterns in different tissues in vitro transient expression by bombarding.

Fifthly, an *Athila*-like retrotransposon was identified in this barley *Mla* region. *Athila* is a new class of retrotransposons with no obvious sequence similarity with the well-characterized *Copia* and *Gypsy* classes. The *Athila* class of retrotransposons has been only previously reported in *Arabidopsis* and is cetromere-orientated (Pelissier et al. 1995; 1996). The published genomic sequences and barley EST database searches indicate that the *Athila*-like elements are highly repetitive in the barley genome and are transcriptionally active (F. Wei and R. P. Wise, unpublished). To get more knowledge on this class of retro elements, we need to investigate the host spectrum, abundance and physical location of this retrotransposon in various species.

Finally, a gene-rich heterochromatic knob, knob1HS1, was characterized in the *Mla* complex. Although two heterochromatic knobs have recently been cloned from the model plant *Arabidopsis*, the mechanism of what and how they function is unknown. Several heterochromatin-related proteins have been identified in yeast (Paszkowski and Whitham 2001). With cloning of knob1HS1 into yeast as an artificial chromosome, we are able to test whether these yeast proteins can function in the barley knob by the immuno-FISH approach (fluorescence in situ hybridization; Brown et al. 1997). Thus, we can get insight into whether the knob and other heterochromatin share the same mechanism.

References

Beck, S., and Trowsdale, J. (2000) The human major histocompatibility complex: lessons from the DNA sequence. Annu. Rev. Genom. Hum. Genet. 1:117-137.

Brown, K.E., Guest, S.S., Smale, S.T., Hahm, K., Merkenschlager, M., and Fisher, A.G. (1997) Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. Cell 91:845-854.

CSHL/WUGSC/PEB *Arabidopsis* sequencing Consortium, (2000) The complete sequence of a heterochromatic island from a higher eukaryote. Cell 100:377-386.

Fransz, P.F., Armstrong, S., de Jong, J.H., Parnell, L.D., van Drunen, C., Dean, C., Zabel, P., Bisseling, T., and Jones, G.H. (2000) Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region. Cell 100:367-376.

Halterman, D., Zhou, F., Wei, F., Wise, R.P., and Schulze-Lefert, P. The MLA6 Coiled-Coil, NBS-LRR protein confers *AvrMla6*-dependent resistance specificity to *Blumeria graminis* f. sp. *hordei* in barley and wheat. Plant J. 25:335-348 (2001).

Lin, X., Kaul, S., Rounsley, S., Shea, T.P., Benito, M.I., et al. (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature 402:761-768.

Paszkowski, J., and Whitham, S.A. (2001) Gene silencing and DNA methylation processes. Curr. Opin. Plant Biol. 4:123-129.

Pelissier, T., Tutois, S., Deragon, J.M., Tourmente, S., Genestier, S., and Picard, G. (1995) Athila, a new retroelement from *Arabidopsis thaliana*. Plant Mol. Biol. 29:441-452.

Pelissier, T., Tutois, S., Tourmente, S., Deragon, J.M., and Picard, G. (1996) DNA regions flanking the major *Arabidopsis thaliana* satellite are principally enriched in Athila retroelement sequences. Genetica97:141-151.

Tabata, S., Kaneko, T., Nakamura, Y., Kotani, H., Kato, T., et al. (2000) Sequence and analysis of chromosome 5 of the plant *Arabidopsis thaliana*. Nature 408:823-826.

Tadauchi, T., Matsumoto, K., Herskowitz, I., and Irie, K. (2001) Post-transcriptional regulation through the HO 3'-UTR by *Mpt5*, a yeast homolog of *Pumilio* and *FBF*. EMBO J. 20:552-561.

Wei, F., Gobelman-Werner, K., Morroll, S.M., Kurth, J., Mao, L., Wing, R., Leister, D., Schulze-Lefert, P., and Wise, R.P. (1999) The *Mla* (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. Genetics 153:1929-1948.

Zhang, P., Chopra, S., and Peterson, T. (2000) A segmental gene duplication generated differentially expressed *myb*-homologous genes in maize. Plant Cell 12:2311-2322.

Zhou, F., Kurth, J., Wei, F., Elliott, C., Vale, G., Yahiaoui, N., Keller, B., Somerville, S., Wise, R.P. and Schulze-Lefert, P. (2001) Cell-autonomous expression of barley *Mla1* confers race-specific resistance to the powdery mildew fungus via a *Rar1*-independent signaling pathway. Plant Cell 13:337-350.

ACKNOWLEDGEMENTS

First of all, I would like to thank my major advisor, Dr. Roger Wise, for his tremendous support, both economically and spiritually during my Ph.D study. He is a good tutor to teach me many lab techniques and to guide my scientific career. His pay-attention-in-detail would greatly benefit my future research.

I also thank my POS members, Drs. Allen Miller, Thomas Baum, Randy Shoemaker and Theodore Bailey, their guidance definitely enrich my research life at Iowa State University and enable me fulfill my dream to be a Ph.D and a scientist. I also express my thanks to all my colleagues at Wise's Lab. They are Karin Gobelman-Werner, Drs. Dennis Halterman, Shaun Morroll, Rich DeScenzo, and Arla Bush, Deqing Pei, GongXin Yu, Carren Dill, and Jennifer Holmes. They all help me in different ways to finish the degree. I definitely need to thank our collaborators, Drs. Rod Wing and Long Mao in Clemson University, Drs. Paul Schulze-Lefert and Fazong Zhou at Sainsbury Laboratory in the John Innes Centre in the U.K. As a team, we finished our dream to clone the barley *Mla* powdery mildew resistance genes.

As always, I thank my wife, Xiaoxia Li for her support and understanding during this long study. Without her, I can not come to my place today. I also feel indebted to my daughters Alexandra and Zoey for not having enough time to play with them. I also thank my Mom and Dad. They always have confidence on me.